EARLY PREDICTION OF PREECLAMPSIA RISK IN PREGNANT WOMEN UTILIZING RANDOM FOREST AND PARTICLE SWARM OPTIMIZATION TECHNIQUES

R. TOPAN ADITYA RAHMAN

SULTAN IDRIS EDUCATION UNIVERSITY

2025

EARLY PREDICTION OF PREECLAMPSIA RISK IN PREGNANT WOMEN UTILIZING RANDOM FOREST AND PARTICLE SWARM OPTIMIZATION TECHNIQUES

R. TOPAN ADITYA RAHMAN

THESIS PRESENTED TO QUALITY FOR DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTING AND META-TECHNOLOGY
SULTAN IDRIS EDUCATION UNIVERSITY

2025

UPSI/IPS-3/BO 32
Pind : 00 m/s: 1/1

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

Sila tanda (√)
Kertas Projek ☐
Sarjana Penyelidikan ☐
Sarjana Penyelidikan dan Kerja Kursus ☐
Doktor Falsafah ☑

INSTITUT PENGAJIAN SISWAZAH

PERAKUAN KEASLIAN PENULISAN

Perakuan ini telah dibuat pada ...12...(hari bulan)...Februari... (bulan) 20..25...

i. Perakuan pelajar :

Saya, R. Topan Aditya Rahman, P20211002731, Faculty of Computing and Meta-Technology (SILA NYATAKAN NAMA PELAJAR, NO. MATRIK DAN FAKULTI) dengan ini mengaku bahawa disertasi/~~tesis~~ yang bertajuk Early Prediction of Preeclampsia Risk in Pregnant Women Utilizing Random Forest and Particle Swarm Optimization Technique adalah hasil kerja saya sendiri. Saya tidak memplagiat dan apa-apa penggunaan mana-mana hasil kerja yang mengandungi hak cipta telah dilakukan secara urusan yang wajar dan bagi maksud yang dibenarkan dan apa-apa petikan, ekstrak, rujukan atau pengeluaran semula daripada atau kepada mana-mana hasil kerja yang mengandungi hak cipta telah dinyatakan dengan sejelasnya dan secukupnya

Tandatangan pelajar

ii. Perakuan Penyelia:

Saya, Assoc. Prof Madya. Dr. Mohammad Modi Bin Lakulu (NAMA PENYELIA) dengan ini mengesahkan bahawa hasil kerja pelajar yang bertajuk Early Prediction of Preeclampsia Risk in Pregnant Women Utilizing Random Forest and Particle Swarm Optimization Technique (TAJUK) dihasilkan oleh pelajar seperti nama di atas, dan telah diserahkan kepada Institut Pengajian Siswazah bagi memenuhi sebahagian/sepenuhnya syarat untuk memperoleh Ijazah Philosophy of Doctoral (SLA NYATAKAN NAMA IJAZAH).

12 Februari 2025
Tarikh

Tandatangan Penyelia

UPSI/IPS-3/BO 31
Pind.: 01 m/s:1/1

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

INSTITUT PENGAJIAN SISWAZAH /
INSTITUTE OF GRADUATE STUDIES

BORANG PENGESAHAN PENYERAHAN TESIS/DISERTASI/LAPORAN KERTAS PROJEK
DECLARATION OF THESIS/DISSERTATION/PROJECT PAPER FORM

Tajuk / *Title*: Early Prediction of Preeclampsia Risk in Pregnant Women Utilizing Random Forest and Particle Swarm Optimization Technique

No. Matrik /*Matric's No.*: P20211002731

Saya / *I* : R. Topan Aditya Rahman

(Nama pelajar / *Student's Name*)

mengaku membenarkan Tesis/Disertasi/Laporan Kertas Projek (Kedoktoran/Sarjana)* ini disimpan di Universiti Pendidikan Sultan Idris (Perpustakaan Tuanku Bainun) dengan syarat-syarat kegunaan seperti berikut:-
acknowledged that Universiti Pendidikan Sultan Idris (Tuanku Bainun Library) reserves the right as follows:-

1. Tesis/Disertasi/Laporan Kertas Projek ini adalah hak milik UPSI.
   The thesis is the property of Universiti Pendidikan Sultan Idris
2. Perpustakaan Tuanku Bainun dibenarkan membuat salinan untuk tujuan rujukan dan penyelidikan.
   Tuanku Bainun Library has the right to make copies for the purpose of reference and research.
3. Perpustakaan dibenarkan membuat salinan Tesis/Disertasi ini sebagai bahan pertukaran antara Institusi Pengajian Tinggi.
   The Library has the right to make copies of the thesis for academic exchange.
4. Sila tandakan (√) bagi pilihan kategori di bawah / *Please tick (√) for category below:-*

☐ **SULIT/*CONFIDENTIAL*** — Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub dalam Akta Rahsia Rasmi 1972. / *Contains confidential information under the Official Secret Act 1972*

☐ **TERHAD/*RESTRICTED*** — Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan ini dijalankan. / *Contains restricted information as specified by the organization where research was done.*

☑ **TIDAK TERHAD / *OPEN ACCESS***

(Tandatangan Pelajar/ Signature)

Tarikh: 12 Februari 2025

Prof. Madya Ts Dr. Muhammad Modi Lakulu
(Tandatangan Penyelia / *Signature of Supervisor*)
& (Nama & Cop Rasmi / *Name & Official Stamp*)
Fakulti Komputeran dan Meta-Teknologi
Universiti Pendidikan Sultan Idris
35900 Tanjong Malim, Perak

Catatan: Jika Tesis/Disertasi ini **SULIT @ TERHAD**, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh laporan ini perlu dikelaskan sebagai **SULIT** dan **TERHAD**.

Notes: If the thesis is CONFIDENTAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction.

ACKNOWLEDGMENT

"In the name of Allah, the Most Gracious, the Most Merciful". Alhamdulillahhirobbil'alamin is expressed by the researcher to Allah SWT for His blessings, mercy, and gifts so that the researcher can complete this research proposal. Sending sholawat and greetings to the Great Prophet Muhammad SAW, his family, and friends; hopefully, his blessings will always be bestowed on all humankind. The researcher also thanks the main supervisor, Prof. Madya Dr. Muhammad Modi Bin Lakulu, and Dr. Ismail Yusuf Panessai for their guidance when the researcher completed this research with great patience and endless encouragement. I also thank the examiner, Dr. Suliana binti Sulaiman, for her guidance and input in the completion of this thesis The researcher also expresses gratitude for the love and affection of his beloved wife and children, who have provided endless encouragement while the resarcher is completing this research. The researcher also thanks my parent and the big family, who has provided material and moral support so that the researcher has the opportunity to pursue Doctoral studies, as well as friends in arms who are always together and help each other in completing this research. The researcher realizes this research would not have been successful without all support.

ABSTRACT

The main purpose of this study is to predict early preeclampsia using machine learning algorithms with Particle Swarm Optimization. Preeclampsia is one of the causes of maternal mortality, in the last two decades, there has been no significant decrease in the incidence of preeclampsia. The magnitude of this problem has an impact on the mother during pregnancy and childbirth. The entire population in this study was pregnant women. The entire population in this study was pregnant women. The samples in this study were 504 pregnant women from the medical records of Ansari Saleh General Hospital Banjarmasin in 2022. The algorithm used in this study using eXtreme Gradient Boosting, Adaptive Boosting, Random Forest, Logistic Regression, and for optimization algorithm using Particle Swarm Optimization. Based on the result, Random Forest was the best model with an accuracy rate of 96.08%. The variables that most influence the incidence of preeclampsia are the history of preeclampsia, a history of hypertension, a history of caesarean section delivery, and a history of diabetes mellitus. This successful evaluation of model development provides implications to help health workers in carrying out pregnancy screening, the trust of expectants in the service quality given surely affects society to utilize technology-based service rather than the conventional one. It recommends developing a prototype application for the early detection of preeclampsia using machine learning technology to assist healthcare professionals in delivering optimal antenatal care and transitioning to technology-based pregnancy monitoring as a proactive measure for pregnant women, ultimately helping to prevent preeclampsia.

RAMALAN AWAL RISIKO PRA-EKLAMPSIA DALAM KALANGAN WANITA HAMIL MENGGUNAKAN TEKNIK HUTAN RAWAK DAN PENGOPTIMUMAN KERUMUNAN ZARAH

ABSTRAK

Tujuan utama kajian ini adalah untuk meramalkan Pra-eklampsia awal menggunakan algoritma pembelajaran mesin, dengan Pengoptimuman Kerumunan Zarah. Pra-eklampsia adalah salah satu punca kematian ibu, dalam dua dekad yang lalu, tiada penurunan ketara dalam kejadian Pra-eklampsia. Isu ini mempunyai kepentingan yang besar dan memberikan impak kepada ibu semasa tahap kehamilan serta proses bersalin. Populasi keseluruhan dalam kajian ini terdiri daripada wanita yang sedang hamil. Sampel dalam kajian ini adalah seramai 504 wanita hamil dari rekod perubatan Ansari Saleh Banjarmasin Hospital pada tahun 2022. Algoritma yang digunakan dalam kajian ini adalah Peningkatan Kecerunan ekstrem, Penggalak Adaptif, Hutan Rawak, Regresi Logistik, dan algoritma pengoptimuman menggunakan Pengoptimuman Kerumunan Zarah. Berdasarkan keputusan tersebut, Hutan Rawak adalah model terbaik dengan kadar ketepatan 96.08%. Faktor yang paling mempengaruhi kejadian Pra-eklampsia adalah sejarah Pra-eklampsia, sejarah hipertensi, sejarah kelahiran melalui caesarean, dan sejarah diabetes mellitus. Penilaian kejayaan pembangunan model ini memberi implikasi untuk membantu petugas kesihatan dalam menjalankan saringan kehamilan, kepercayaan jangkaan terhadap kualiti perkhidmatan yang diberikan sudah pasti mempengaruhi masyarakat untuk menggunakan perkhidmatan berasaskan teknologi berbanding konvensional. Dicadangkan pembangunan aplikasi prototaip untuk pengesanan awal Pra-eklampsia menggunakan teknologi pembelajaran mesin bagi membantu profesional kesihatan dalam memberikan penjagaan antenatal yang optimum, serta beralih kepada pemantauan kehamilan berasaskan teknologi sebagai langkah proaktif untuk wanita hamil, seterusnya membantu mencegah Pra-eklampsia.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	ACCURACY
ACOG	AMERICAN COLLAGE OF OBSTSETRICIANS AND GYNECILOGISTS
ADHD	ATTENTION DEFICIT HYPERACTIVITY DISORDER
AI	ARTIFICIAL INTELLIGENCE WHO WORLD
ANC	ANTENATAL CARE
ARL	ASSOCIATION RULE LEARNING
AUC	AREA UNDER CURVE
BEDROC	BOLTZMANN ENHANCED DISCRIMINATION OF ROC
BMI	BODY MASS INDEX
CART	CLASSIFICATION AND REGRESSION TREES
CASP	CRITICAL APPRAISAL SKILL PROGRAMME
CI	COEFICIENCE INTERVAL
CNN	CONVOLUTION NEURAL NETWORK
CRH	CORTICOTROPIC RELEASING HORMON
CV	CROSS VALIDATION
CVD	CARDIOVASCULAR DISEASE

DARPA	DEFENSE ADVANCED RESEARCH PROJECT AGENCY
DBM	DEEP BLOTZMAN MCHINE
DIC	DISSEMINATED INTRAVASCULAR COAGULATION
DNA	DEOXYRIBONUCLEIC ACID
DL	DEEP LEARNING
DOC	DEOXYCORTICOSTERONE
DV	DECIDUAL VASCULOPATHY
ECM	EXTRACELLULAR MATRIX
EHR	ELECTRONIC HEALTH RECORD
FN	FALSE NEGATIVE
FP	FALSE POSITIVE
GB	GRADIENT BOOSTING
GFR	GLOMERULAR FILTRATION RATE
GH	GESTATIONAL HYPERTENSION
HB	HAEMOGLOBIN
HCG	HUMAN CHORIONIC GONADOTROPIN
HELLP	HEMOLYSIS ELEVATED LIVER ENZYMES AND LOW PLATELETS
HLA-G	HUMAN LEUKOCYTE ANTIGEN PROTEIN G
HPD	HYPERTENSION PREGNANCY DISORDER
ICU	INTENSIVE CARE UNIT
ISO	INTERNATIONAL ORGANIZATION FOR STANDARDIZATION

IUFD	INTRA UTERINE FETAL DEATH
IUGR	INTRAUTERINE GROWTH RESTRICTION
KG/M	KILOGRAM IN METER
KNN	K-NEAREST NEIGHBOR
LBF	LOW BACK PAIN
LDA	LINIAR DISCRIMONAT ANALYSIS
LR	LOGISTIC REGRESSION
LWF	LOCALLY WEIGHTED LEARNING
LVQ	LEARNING VECTOR QUANTIZATION
MCC	CORRELATION COEFFICENT MATTHEW
MDA	MIXED DISCRIMINANT ANALYSIS
ML	MACHINE LEARNING
MMHG	MILIMETER HYDRARGYRUM
MMR	MATERNAL MORTALITY RATE
NICE	NATIONAL INSTITUTE FOR HEALTH AND EXCELLENCE
NK	NATURAL KILLER
NLP	NATURAL LANGUAGE PROCESSING
PCA	PRINCIPAL COMPONENT ANALYSIS
PCR	PRICIPLE COMPONENT REGRESSION
PCR	PROTEIN TO CREATININE RATIO
PSO	PARTICLE SWARM OPTIMIZATION
PE	PREECLAMPSIA
PlGF	PLACENTAL GROWTH FACTOR
PIH	PREGNANCY INDUCED HYPERTENSION

PRE	PRECISSION
RF	RANDOM FOREST
RNN	RECURRENT NEURAL NETWORK
ROC	RECEIVER OPERATING CHARACTERISTIC
RTMS	REPETITIVE TRANSCRANIAL MAGNESIUM STIMULATION
SGOT	SERUM GLUTAMIC OXALOACETIC TRANSAMINASE
SGPT	SERUM GLUTAMIC OXALOACETIC TRANSAMINASE
SENS	SENSITIVITY
SPEC	SPECIFICITY
SVM	SUPPORT VECTOR MACHINE
TN	TRUE NEGATIVE
TP	TRUE POSITIVE
VEGF	VASCULAR ENDOTHELIAL GROWTH FACTOR
WHO	WORLD HEALTH ORGANIZATION

LIST OF APPENDIXES

CHAPTER 1

INTRODUCTION

1.1. Overview

The health of pregnant women is essential in the development of the fetus, so it needs an optimal monitoring. Monitoring during pregnancy currently can be done through technology-based examinations such as ultrasound which can assist health workers in making decisions during pregnancy (Abuelezz et al., 2022). Preeclampsia is a leading cause of maternal and fetal morbidity and mortality in the general population (Udenze et al., 2017). Over the past ten years, artificial intelligence technology has significantly been used in the industrial and health sectors. Most recently, AI-based technologies have been used to diagnose pregnancy. AI can be applied in health care for modeling, diagnosing, early detecting, and monitoring due to the disease's varied gene expressions. Such issues can be solved using machine learning technologies (Sufriyana et al., 2020a).

Chapter 1 is presented to introduce the idea of this research. First, Section 1.2 Research Background. Second is Section 1.3, the problem statement stated by the researcher. Third is section 1.4 explains the Research Questions. Fourth, Section 1.5 Research Objectives. Then, Section 1.6 Research Scopes. Then, Section 1.7, The importance of Research. Then, section 1.8 Research Organization. Then, section 1.9 Operation Definition. Then, section 1.10 explains the summary of the research.

1.2. Research Background

According to the World Health Organization (WHO), more than eight million women per year experience pregnancy complications and more than half a million of them die, of which 99% occur in developing countries. The death rate due to complications of pregnancy and childbirth in developed countries is 1 in 5000 women, which is much lower than in developing countries, specifically 1 in 11 women die due to complications of pregnancy and childbirth. Over 800 women die every day because of preventable causes related to pregnancy's inherent dangers around the world (Bertini et al., 2022). Preeclampsia is a leading cause of maternal and fetal morbidity and mortality in the general population. The three main causes of maternal death are bleeding (30%), hypertension in pregnancy (25%), and infection (12%) (Udenze et al., 2017) (Munchel, 2020).

High blood pressure (hypertension), fluid retention (edema), and an increased protein rate in the urine are all symptoms of preeclampsia (proteinuria). Because this evidence is not visible at the outset of pregnancy, preeclampsia may be difficult to detect (Tahir et al., 2018). Preeclampsia is a pregnancy-related hypertension condition

(HPD), and it affects 2%–8% of all pregnancies; thus, it becomes a leading cause of maternal and perinatal morbidity as well as mortality (Ives et al., 2020a). Preeclampsia is a kind of pregnancy-induced hypertension (PIH), the leading cause of maternal death in the United States (Tejera et al., 2021). Although the etiology and pathogenesis of preeclampsia are unknown, it is known that childbirth is the only cure; consequently, a false-positive test predicting preeclampsia may result in unduly early delivery. Premature and low birth weight newborns result from this, leading to more inefficient use of the neonatal Intensive Care Unit (ICU). The predictive model of preeclampsia developed in prior studies has four flaws: (1) no significant predictor exists for all preeclampsia subtypes; (2) biased predictive performance; (3) low precision or positive predictive value; and (4) the need for high resource settings to develop predictive models (Sufriyana et al., 2020a).

Many studies are used to define a woman as a woman at high risk of developing preeclampsia. In this study, the variables as risk factors include maternal age, education, parity, pregnancy interval, hypertension, hemoglobin, body mass index, history of preeclampsia, and gestational diabetes mellitus (Bartsch et al., 2016) (Chu et al., 2020) (Lewandowska et al., 2020) (Liu et al., 2021). The incidence of preeclampsia increased two until four-fold in 2017 in women living at high altitudes, implying hypoxia may be a contributing factor to the occurrence of preeclampsia. Hypertension and proteinuria can be caused by constriction of uterine blood flow in pregnant women in primates and other mammals. These observations suggest placental ischemia may be an early event.

In Indonesia, there are 3 (three) leading causes of maternal death, namely bleeding (28%), preeclampsia/eclampsia (24%), and infection (11%). According to the

National Guidelines for Medical Services, the incidence of preeclampsia in Indonesia is 128,273/year or 5,3% (Kementerian Kesehatan RI, 2016). The latest survey data shows Indonesia's Maternal Mortality Rate (MMR) is 305/100,000 live births (Kementerian Kesehatan RI, 2021). The trend in the last two decades has not seen a significant decrease in the incidence of preeclampsia, in contrast to the incidence of infection, which has decreased according to the development of antibiotic findings. Preeclampsia is a specific condition in pregnancy characterized by placental dysfunction and maternal response to systemic inflammation with endothelial activation and coagulation. The diagnosis of preeclampsia is made based on the presence of specific hypertension caused by pregnancy accompanied by disorders of other organ systems at a gestational age above 20 weeks. Preeclampsia was previously used to be defined by the presence of hypertension and proteinurine that just occurred in pregnancy (new onset hypertension with proteinuria). Even though these two criteria are still the classic definition of preeclampsia, several other women show hypertension accompanied by other multisystem disorders. It indicates the presence of severe preeclampsia, even if the patient does not experience urine proteinuria.

The maternal mortality rate in South Kalimantan tends to fluctuate over the last 3 years. In 2018 the maternal mortality rate reached 135/100,000 live births. There was a decrease in 2019, which reached 92 cases, and an increase again in 2020 which reached 135 cases. The cause of maternal deaths is primarily due to bleeding and complications of pregnancy and childbirth, namely preeclampsia/eclampsia (Dinas Kesehatan Provinsi, 2021).

Dr. H. Moch Ansari Saleh General Hospital Banjarmasin as a hospital owned by the local government and as a referral hospital, is a hospital that has a high incidence of preeclampsia. According to a preliminary study conducted during the last 3 (three) years, the number of cases has been increasing. The incidence of preeclampsia in 2020 was 178 (5,92%) of 3,007 mothers who gave birth; in 2021 there were 145 (5,17%) cases of 2,804 women who gave birth and increased again in 2022 at 168 (9,27%) of 1,813 cases of mothers giving birth. This percentage is still high because the government's target through the Ministry of Health is that the maternal mortality rate is less than 5% per year (Erni Setiawati, 2020) (Dinas Kesehatan Provinsi Kalimantan Selatan, 2022) (Menteri Kesehatan RI, 2020). These data indicate the high incidence of preeclampsia in dr. H. Moch Ansari Saleh General Hospital Banjarmasin every year.

Over the past ten years, artificial intelligence technology has significantly been used in the industrial and health sectors. Most recently, AI-based technologies have been used to diagnose pregnancy (Abuelezz et al., 2022). Artificial Intelligence (AI) based technologies have been hailed as a potential tool for assessing disparate data sources. Over 75% of research AI-based technology has been lauded as a potential tool in predicting pregnancy or pregnancy disorders (Oprescu et al., 2020). It has the potential to assist the process of decision-making and improve medical care. AI can be applied in health care for modelling, diagnosing, early detection, and monitoring. Another potential difficulty is the noisy class of outcomes due to the disease's varied gene expressions. Such issues can be solved using machine learning technologies (Sufriyana et al., 2020b).

Machine learning is well-suited for predictive modeling of pregnancy outcomes (Espinosa et al., 2021). Machine Learning (ML), also known as supervised, semi-supervised, unsupervised, or reinforcement learning, is a subset of artificial intelligence that involves the use of algorithms and computer models to achieve a certain goal. In decision-making settings, machine-learning algorithms are commonly utilized to achieve higher predictive accuracy than traditional generalized linear models (Liu et al., 2021). Furthermore, deep learning is a machine learning technique that employs neural networks, similar to neurons in the human brain, to extract many levels of data representation from a given input in order to solve a problem (Ramakrishnan et al., 2021).

The optimization algorithm used in this study is Particle Swarm Optimization (PSO). Given the difficulty of predicting preeclampsia, data mining can play an important role in accurately predicting preeclampsia. This study uses a machine learning approach with particle swarm optimization (PSO) and Random Forest, Adaboost, XGBoost, as well as Logistic Regression. The model is a new approach proposed to predict preelampsia (Asadi et al., 2021). PSO as a possible optimization tool can be a good alternative in optimizing decision tree-based classification rules (Nilawati & Achyani, 2019). PSO is one of the optimization algorithms used in decision-making to determine process parameters that produce the optimum response value (Sateria et al., 2018). PSO is also an effective algorithm for solving combinatorial problems (Rizki, 2021).

According to the statement of one of the research experts, who is also a specialist in Obstetrics and Gynecology, the pathogenesis of preeclampsia still needs

to be explained (Hadiati, 2017). To date, early detection of preeclampsia has been carried out conventionally through antenatal care examinations, so artificial technology support is needed to detect preeclampsia earlier. (Taufik, 2022). Artificial Intelligence-based technology is expected to be able to improve health services, especially during pregnancy. Therefore, it is hoped that further research is highly recommended to find the most effective model for diagnosing pregnancy, specifically preeclampsia (Abuelezz et al., 2022).

As times and technology develop, the machine learning algorithms used are growing and getting more sophisticated as they continue. From several research results, the use of machine learning algorithms in making predictions in cases of preeclampsia also shows a higher level of accuracy. According to the research from Li et al. (2021), XGBoost is the best model with accuracy of 0,920 and precision of 0.447, recall of 0.789, f1_score of 0.571, and auROC of 0.955. Liu et al. (2021) show the AdaBoost model performed the best with an accuracy of 0,951 and an AUC 0,701. Furthermore, Kartik et al. (2021) show the logistic regression model had the best accuracy (0.94, CI: 0.92-0.96). Sufriyana et al. (2020) also show the Random Forest model had the best algorithm with an accuracy of 0,926 and a sensitivity of 90.7.

This study specifically discusses developing a new model for early risk detection of preeclampsia using a random forest algorithm with particle swarm optimization in pregnant women, which is then classified as a risk factor. The study on preeclampsia using machine learning carried out in Indonesia, particularly Banjarmasin, still has a very high limitation.

1.3. Problem Statement

Preeclampsia is the second most common cause of maternal death after bleeding. WHO estimates that the cases of preeclampsia are seven times higher in developing countries than in developed countries. In the last two decades, there has been no significant decrease in the incidence of preeclampsia. The magnitude of this problem is not only because preeclampsia has an impact on the mother during pregnancy and childbirth, but it also causes postpartum problems due to endothelial dysfunction in various organs, such as the risk of cardiometabolic diseases and other complications. Thus, preeclampsia becomes a serious and complex medical problem (Rana et al., 2019). There is no active screening evaluation for preeclampsia, so efforts to prevent preeclampsia are not optimal, which can lead to increased morbidity and mortality. Therefore, there is an urgency for recommendations based on scientific evidence to assist practitioners in diagnosing, evaluating, and managing preeclampsia.

Up to now, there have been various biomarker findings that can be used to predict the incidence of preeclampsia, but no single test has high sensitivity and specificity. This occurs because the features used are mostly related to characteristics. A series of checkups using more features is urgently needed through technological assistance to screen the risk of preeclampsia in pregnant women from the start of their pregnancy, so health practitioners can identify risk factors for preeclampsia and control them as a form of primary prevention. (Menteri Kesehatan RI, 2017).

Pregnancy screening at dr. H. Moch Ansari Saleh General Hospital is currently still being carried out conventionally, which is through antenatal checkups. However,

this is still not optimal and has the potential to produce inaccurate results, considering that cases of preeclampsia still occur and are increasing every year (Taufik, 2022). According to a statement by a research expert who is also a specialist in Obstetrics and Gynecology, most health workers in dealing with pregnancy complications still focus on treatment, which will definitely cost more than prevention in terms of funding (Hadiati, 2017).

Based on a preliminary study conducted from the latest data in 2022, there are 168 cases of preeclampsia at dr. H. Moch Ansari Saleh General Hospital Banjarmasin as the first referral hospital from 1813 birth. Out of the 168 cases of preeclampsia, 60 patients (33.7%) experienced further effects, such as 32 patients experienced bleeding (17.9%), 7 people experienced IUFD (3.9%), 19 patients experienced LBW (10.8%), and the maternal mortality rate is 2 patients (1.1%). Hence, this needs to be a concern for the health sector, especially the early prevention in Banjarmasin so that cases of preeclampsia can be minimized. In consequence, it is necessary to prevent preeclampsia from the first trimester, so pregnant women who are indicated to have preeclampsia can be prevented through pregnancy monitoring until they give birth.

Despite of the fact that there have been several research results regarding the prediction of the incidence of preeclampsia using a machine-learning approach, in reality, the cases of preeclampsia are still one of the highest causes of death for pregnant women. Therefore, it is hoped that further research is highly recommended to find the most effective model for diagnosing pregnancy, especially preeclampsia (Abuelezz et al., 2022). Machine-learning is well-suited for the predictive modeling of pregnancy outcomes (Espinosa et al., 2021). Furthermore, there has never been any research on

machine-learning-based preeclampsia in the dr. H. Moch Ansari Saleh General Hospital Banjarmasin. The following is the result of developing a machine-learning model for detecting preeclampsia. According to the results of research from Li, et al (2019), XGBoost is the best model with an accuracy of 0.920 and precision = 0.447, recall = 0.789, f1_score = 0.571, and auROC = 0.955. Liu, et al (2021) showed that the AdaBoost model performed best with an accuracy of 0.951 and an AUC of 0.701. Kartik, et al (2021) showed that the logistic regression model has the best accuracy with 0.94 CI: 0.92-0.96. Furthermore, Sufriyana, et al (2020) also showed that the RF model has the best algorithm with an accuracy of 0.926 and a sensitivity of 90.7.

While machine learning holds great potential for improving the detection and management of preeclampsia, several significant challenges must be addressed. These include issues with data quality and availability, difficulties in selecting and extracting relevant features, the need for interpretable models that can gain clinical trust, concerns over model generalizability and bias, and the integration of these models into existing clinical workflows. Overcoming these obstacles will require improvements in data collection and processing and the development of models that are not only accurate but also transparent and adaptable to diverse patient conditions histories.

According to the problem statement above, the main purpose of this study is to early prediction of preeclampsia risk using Random Forest algorithm and Particle Swarm Optimization technique, such as to identify the features that influence the incidence of Preeclampsia, to design a process model for early prediction of preeclampsia using machine learning algorithms, and to evaluate the effectiveness of process model for early prediction of preeclampsia in pregnant women. This research

is essential because initial diagnosis that still uses conventional methods is one of the crucial things that need to be developed using machine learning-based information technology.

1.4. Research Objectives

The aim of this research is early detection of preeclampsia by using machine learning with multiple algorithms based on the causes of preeclampsia. This study identifies the factors that cause preeclampsia based on the medical record in Ansari Saleh General Hospital, Banjarmasin. The results of this study are expected to contribute positively to the government in Indonesia, especially in anticipating the occurrence of maternal deaths caused by preeclampsia.

The objectives of this research are:

1. To identify the features that influence the incidence of Preeclampsia.
2. To design a process model for early prediction of preeclampsia using Random Forest Algorithm with particle swarm optimization.
3. To evaluate the effectiveness of process model for early prediction of preeclampsia in pregnant women using Random Forest Algorithm with Particle Swarm Optimization.

1.5. Research Questions

The issues addressed in this study were described in the preceding part, which led to the presentation of the research questions. The questions that would be answered in this

question were regarding the factors that cause preeclampsia and how to early predict preeclampsia using Random Forest algorithm with Particle Swarm Optimization. Specifically, this research was conducted to answer the following questions:

1. What features have influenced the most on the incidence of preeclampsia in pregnant women?
2. How to design a process model for early prediction of preeclampsia in pregnant women using Random Forest algorithm with particle swarm optimization?
3. How to evaluate the process model for early prediction of preeclampsia in pregnant women using Random Forest algorithm with particle swarm optimization?

The research objective and research questions can be seen in Table 1.1 as follows:

Table 1.1

Relation of objectives and research questions

No	Research Objectives	Research Questions
1	To identify the features that influence the incidence of Preeclampsia in pregnant women	What features have influenced the most on the incidence of preeclampsia in pregnant women
2	To design process model for early prediction of preeclampsia using Random Forest algorithm with particle swarm optimization	How to design process model for early prediction of preeclampsia in pregnant women using Random Forest algorithm with particle swarm optimization
3	To evaluate the effectiveness of process model for early prediction of preeclampsia in pregnant women using Random Forest algorithm with particle swarm optimization	How to evaluate the process model for early prediction of preeclampsia in pregnant women using Random Forest algorithm with particle swarm optimization

1.6. Research Hypotheses

A hypothesis is generally defined as a temporary answer (alleged) to a research problem. Hypotheses are only prepared in inferential research, namely research with a quantitative approach that aims to test. Testing a hypothesis always uses inferential statistical analysis techniques. The formulation of a hypothesis has requirements or characteristics that must be met by the researcher. Some of the characteristics of hypothesis formulation are:

a. Null hypothesis (Ho) is a statement that states there is no relationship or difference between the variables studied. In some research or studies, this null hypothesis also says that the expected effect or influence is equal to zero. In scientific research, the null hypothesis is often symbolized by Ho.
b. Alternative hypothesis (Ha) is another answer to a research question that states there is a relationship or difference between the variables studied. The alternative hypothesis is a complement to the null hypothesis. The null hypothesis and alternative hypothesis are exhaustive, meaning they cover all possible outcomes. They are also mutually exclusive, meaning only one is correct at a time.

The hypothesis in this study was to identify factors that have a statistically significant influence on the incidence of preeclampsia, carried out by analyzing using the chi square test. The hypotheses in this research are:

1. Hypotheses 1: “Maternal age has a significant influence on the incidence of preeclampsia”.
2. Hypotheses 2: “Maternal Profession has a significant influence on the incidence of preeclampsia”.

3. Hypotheses 3: “Maternal education has a significant influence on the incidence of preeclampsia”.
4. Hypotheses 4: “Gravidity has a significant influence on the incidence of preeclampsia”.
5. Hypotheses 5: “Parity has a significant influence on the incidence of preeclampsia”.
6. Hypotheses 6: “History of Abortion has a significant influence on the incidence of preeclampsia”.
7. Hypotheses 7: “History of Caesarean Section has a significant influence on the incidence of preeclampsia”.
8. Hypotheses 8: “Antenatal care has a significant influence on the incidence of preeclampsia”.
9. Hypotheses 9: “Pregnancy Interval has a significant influence on the incidence of preeclampsia”.
10. Hypotheses 10: “Hemoglobin has a significant influence on the incidence of preeclampsia”.
11. Hypotheses 11: “Body Mass Index has a significant influence on the incidence of preeclampsia”.
12. Hypotheses 12: “History of Hypertension has a significant influence on the incidence of preeclampsia”.
13. Hypotheses 13: “History of Preeclampsia has a significant influence on the incidence of preeclampsia”.
14. Hypotheses 14: “History of Diabetes Mellitus has a significant influence on the incidence of preeclampsia”.

1.7. Research Scopes

Preeclampsia is a pregnancy complication characterized by hypertension, leg swelling, and increased urine protein. The focus of this study is the early detection of preeclampsia. Research methodology used several algorithms, which were eXtreme Gradient Boosting (XGBoost), AdaBoost, Logistic Regression (LR), and Random Forest (RF). The Optimization algorithm used Particle Swarm Optimization (PSO) and Validation using Cross Validation. The data collected were medical records for pregnant women at dr. Moch. Ansari Saleh General Hospital Banjarmasin in 2022.

1.8. The importance of research

According to the background above, the maternal mortality rate in Indonesia is still high, especially in Banjarmasin, and preeclampsia is one of the factors. The main obstacle is limited screening method in first trimester to early predict preeclampsia. Therefore, it is necessary to take preventive measures to reduce maternal mortality. Thus, the results of this research may contribute to:

1. Artificial Intelligence
    a. Prioritize and recommend Machine Learning applications in the diagnosis/management of preeclampsia in future investigations.
    b. This research is expected to be a tool for the early detection of the prevention of preeclampsia in pregnant women.

2. Health

   a. Know the factors that cause the incidence of preeclampsia with machine learning approach with particle swarm optimization.

   b. The Indonesian government, especially the city of Banjarmasin, can use the method with the most accurate performance as a tool to predict the incidence of preeclampsia in pregnant women.

   c. As an effort by health workers to provide optimal antenatal care so that the incidence of preeclampsia can be prevented.

1.9. Research organization

There are five chapters in this study. The following are summaries of these chapters:

The backdrop of the research, as well as the research problem, are presented in: Chapter 1: The objective of the study and the scope of the research are also stated in this chapter. The main menu on Chapter 1 such as overview, research background, problem statement, research questions, research objective, research scope, hypotheses, importance of research, research organization, operational definition, and summary.

Chapter 2: Includes a review of the literature, a search of journal databases using queries to find papers that are relevant to the study questions and taxonomy research. The topic of research problems, namely preeclampsia, is also discussed in this chapter, including definitions and factors that cause eclampsia, all of which are merged with machine learning. Aside from that, this chapter includes the research's theoretical framework.

Chapter 3: Gives a thorough explanation or overview of the research methodology. Main menu in chapter three is overview, research methodology, data collection procedure, algorithm process model in early prediction of preeclampsia for pregnant, performance evaluation metrics, and summary.

Chapter 4: Is devoted to assessing and presenting the findings of the study, development and data analysis. The main menu in chapter four is an overview, data processing, findings, correlation attribute with preeclampsia, split data, performance model, evaluation model, features of importance, simulation predicting preeclampsia, final process model, and summary.

Chapter 5: Contains the conclusion and future research of the thesis. The main menu in chapter five is an overview, answering of research objective, contribution of research, research strengths and limitations, recommendations for future research, and conclusions.

1.10. Operational Definition

The variables in the study were based on the results of previous research studies through a literature review. Table 1.2 defines the operational definitions for this research.

Table 1.2

Operational Definition for this Research is:

No		Operational Definition
1	Age	The age of pregnant women can be defined as the length of life of pregnant women as measured from the last birthday according to medical records. (Han et al., 2019) (Nsugbe, 2021)

(continue)

Table 1.2 *(continued)*

2	Profession	Activities pregnant women to obtain rewards or wages. (Khayati & Veftisia, 2018)
3	Education	Education for pregnant women can be defined as formal education that has been taken up to obtain a certificate according to medical records. (Liu et al., 2021) (Manoochehri et al., 2021) (C. R. Jung et al., 2021)
4	Gravidity	One of the factors that influences the incidence of preeclampsia is gravida status, where primigravida mothers are at greater risk of experiencing preeclampsia than multigravida
5	Parity	Parity is the number of births that results in live or dead babies. Parity was significantly associated with all or some of the complications of preeclampsia during pregnancy (Lisonkova et al., 2021)
6	History of abortion	This is due to the inhibition of immature antibody formation (E. Purwanti et al., 2019).
7	History of caesarian section	Uterine involution after Caesarian Section (CS) is slow compared with uterine involution in vaginal delivery (Suryawinata et al., 2019).
8	Antenatal care	Antenatal Care (ANC) in normal pregnancy at least 6 (six) times (Kementerian Kesehatan RI, 2020).
9	Pregnancy Interval	Pregnancy interval can be defined as the distance between the last delivery and the next pregnancy experienced by pregnant women according to medical records. (Lewandowska et al., 2020)
10	Hemoglobin	Hemoglobin is defined as a decrease in red blood cells during pregnancy <11 g/dl in pregnant women according to the diagnosis recorded in the medical record. (Sandström et al., 2019)
11	Body Mass Index	Body Mass Index can be defined as a measurement used to determine healthy and unhealthy weight groups. Weight gain during pregnancy > 30 kg/m2 according to the

(continue)

Table 1.2 *(continued)*

		diagnosis recorded in the medical record. (Sufriyana et al., 2020b) (Lewandowska, 2021)
12	History of Hypertension	A condition in which the blood pressure of pregnant women is above 140/90 mmHg according to the diagnosis recorded in the medical record. (Ives et al., 2020a) (Sandström et al., 2019)
13	History of Preeclampsia	Having a history of increased blood pressure, edema, and proteinuria in a previous pregnancy according to the diagnosis recorded in the medical record. (Tahir et al., 2018) (Sandström et al., 2019) (Yoffe et al., 2018)
14	History of Diabetes Mellitus	Impaired carbohydrate tolerance caused blood sugar levels to rise and was first discovered during pregnancy according to the diagnosis recorded in the medical record. (Sufriyana et al., 2020a) (Clymer et al., 2020)
15	Preeclampsia	Preeclampsia can be defined as pregnancy condition of pregnant women which is characterized by the onset of hypertension, edema, and proteinuria. (Bartsch et al., 2016)
16	Machine Learning	Tools used to predict the incidence of preeclampsia. (Li et al., 2021)
17	Xtreme Gradient Boosting	One part of Machine learning algorithm used in predicting preeclampsia. (Jhee et al., 2019)
18	AdaBoost	The essence of the AdaBoost algorithm is to give more weight to observations that are not right (weak classification) (Roihan et al., 2020).
19	Logistic Regression	Logistic regression is a machine learning algorithm that functions to solve binary classification tasks by making predictions, events, possible outcomes, or observations. (Kanade, 2022).
20	Random Forest (RF)	RF is a part of an effective supervised machine learning algorithm (Chen & Guestrin, 2016).

(continue)

Table 1.2 *(continued)*

21	Particle Swarm Optimization (PSO)	As a possible optimization tool, it can be a good alternative in optimizing decision tree-based classification rules (Nilawati & Achyani, 2019)
22	Cross Validation	To assess the performance measure and generalization ability of a model using cross validation with 10k-Fold Cross Validation

1.11. Summary

The first chapter is an introduction, research background and problem statement, which focuses on models, theories, and previous research on the prediction of preeclampsia. This chapter explains the formulation of the problem to be solved, research objectives, research questions, and research scope, as well as the importance of the research. In addition, the machine learning approach discussed in this chapter can be used as a basis for predicting preeclampsia. ML has a promising ability to predict cases of pregnancy, especially preeclampsia. ML algorithms that have the potential for early detection of preeclampsia are eXtreme Gradient Boosting, AdaBoost, Logistic Regression, and Random Forest (RF), as well as Particle Swarm Optimization (PSO) and Cross Validation for the evaluate algorithm to show the best model.