NELSON MANDELA UNIVERSITY

A Model to Predict the Development of Preeclampsia in South African Women

Nathan Smith

Submitted in fulfilment of the requirements for the degree of Master of Science in the Faculty of Science at the Nelson Mandela University

Supervisor: **Prof Janet Wesson**

Table of Contents

Acknowledgements

Firstly, I would like to thank God for giving me this opportunity, which many do not get. Thank you for giving me the strength and determination to finish this research. I am forever blessed.

I would also like to thank my supervisor, Prof Janet Wesson, for her invaluable guidance and continuous support for the duration of this research. I appreciate everything you have done for me.

A special thank you goes to my wife, who has believed in me more than anyone. I am beyond lucky to have someone always supporting and encouraging me through the ups and downs.

I would also like to thank those that played a part in assisting me with the completion of this dissertation:

- Nelson Mandela Telkom Centre of Excellence in providing research funding and equipment
- The Department of Computing Sciences for enabling me to conduct my research through the resources they provided
- My family and friends, whose continued support and confidence helped me complete this work

Finally, I would like to thank Dr Venter for participating in this study. Thank you for giving me a chance when no one else would. I am incredibly grateful.

Summary

Preeclampsia is the new onset of hypertension and is one of the leading causes of maternal mortality in South Africa and the world. Preeclampsia is usually diagnosed after 20 weeks' gestation. Due to South Africa's poor level of antenatal care, the prediction of pregnant women at risk of developing preeclampsia can be an essential component of improving the level of antenatal. This study used an antenatal care dataset from a South African obstetrician. A review of the literature and existing systems was conducted to identify the eight risk factors. These risk factors are systolic blood pressure, diastolic blood pressure, maternal age, body mass index, diabetes status, hypertension history, nulliparity, and maternal disease. This study used antenatal care datasets from a South African obstetrician. Two models were developed that could accurately predict the development of preeclampsia, one before 16 weeks' gestation and the other within three check-ups. The model was evaluated using five evaluation metrics: classification accuracy, area under the curve, precision, recall and F-Score. The results of this study show a promising future for the use of machine learning models in health care. To the researcher's knowledge, this model is the first machine learning model for predicting preeclampsia using a South African dataset. Future work will revolve around validating the model on data collected from field studies in hospitals and clinics around South Africa.

Keywords: Pregnancy, Preeclampsia, Stillbirth, Machine Learning Models, Prediction, Supervised Learning

Declaration

I, Nathan Smith, hereby declare that the dissertation "**A Model to Predict the Development of Preeclampsia in South African Women",** for the degree of Master of Science in Computer Science and Information Systems**,** is my own independent work. All sources used or quoted have been indicated and acknowledged by means of complete references. This dissertation has not been previously submitted for assessment to any other university or towards the completion of any other qualification.

N Smith

Nathan Smith

List of figures

List of tables

List of equations

Glossary

Abbreviation	Full Definition
LMICs	Lower- and Middle-income counties
NMU	Nelson Mandela University
WHO	World Health Organisation
DSR	Design Science Research
DSRM	Design Science Research Methodology
IT	Information Technology
IS	Information Systems
BMI	Body Mass Index
BP	Blood Pressure
SBP	Systolic Blood Pressure
DBP	Diastolic Blood Pressure
EHR	Electronic Health Records
USA	United States of America
GH	Gestational Hypertension
HDPs	Hypertension Disorders of Pregnancy
EDA	Exploratory Data Analysis
CV	Cross Validation

LR	Logistic Regression
SVM	Support Vector Machine
NB	Naïve Bayes
KNN	K th Nearest Neighbour
DT	Decision Tree
RF	Random Forest
GBC	Gradient Boosted Classifier
SGDC	Stochastic Gradient Descent Classifier

Chapter 1. Introduction

1.1 Background

In 2015, the World Health Organisation (WHO) reported that approximately 303 000 women die annually due to pregnancy complications. WHO also estimated that approximately 2.6 million third-trimester stillbirths occur annually, with 98 per cent occurring in low- and middle-income countries (LMICs) (World Health Organisation, 2016). Forty-one per cent of these 2.6 million stillbirths occurred in Africa. In 2015, Statistics South Africa reported the number of registered stillbirths was approximately 13 681. However, these are only the reported stillbirths; the actual number is potentially much higher. A stillbirth's effect goes far beyond the loss of life. Maternal depression, the financial cost to parents, and economic costs to society are only a few of the overlooked effects of a stillbirth (De Bernis, Kinney, & Delany, 2016). An article by the Centres for Disease Control and Prevention describes the story of a mother who had recently experienced a stillbirth. The following quote from the mother highlights the underlying problem that leads to many stillbirths,

"I wish I had known stillbirth is as common as it is. I wish I had known that having multiple symptoms of preeclampsia was a big deal. I wish my providers had explained this to me or had been more concerned. But most of all, I wish more people talked about stillbirth. Stillbirth affects 1 in 160 pregnancies. That's an enormous number. Stillbirth is so much more than a statistic; it's a life-changing experience. My son should be here, and he isn't. And that will affect me for the rest of my life".

Current understanding of the causes of stillbirths in LMICs is based on unvalidated verbal autopsy data and registration data, which is estimated to record less than five per cent of all stillbirths. Reinebrant et al. (2018) performed a systematic review of stillbirths between 2009 and 2016. This review highlighted some significant factors associated with the causes of foetal death, such as the low-quality data, reliance on data from verbal autopsies and administrative data, and inconsistent use of

classification systems. The lack of antenatal care is also a significant factor in stillbirths and maternal deaths. Almost 50 per cent of women in LMICs do not receive adequate antenatal care (Finlayson & Downe, 2013).

Recently, machine learning models have made considerable advances in healthcare. Machine learning is broadly defined as computational methods that can learn and adapt without specific instructions. Machine learning does this by analysing large amounts of information, known as datasets (Mohri, Rostamizadeh, & Talwalkar, 2018). The following three examples emphasise the necessity of integrating machine learning in the health sector:

- Google's Deep Learning program for detecting breast cancer achieved an accuracy of 89 per cent (Bresnick, 2017);
- A machine learning model that can predict a patient's death achieved an accuracy of 95 per cent (Burgen, 2018); and
- The use of machine learning during the COVID-19 pandemic (An et al., 2020).

The potential for machine learning is limitless as it continues transforming most industries.

1.2 Problem Statement

The following problem statement further defines the problem this research will aim to resolve:

Preeclampsia is a treatable condition that affects many women. In South Africa, many women do not have adequate information on their pregnancy status to know when they should seek medical help, possibly resulting in a stillbirth.

1.3 Aim of Research

The main goal of this research is:

To create a model that can predict if a pregnant woman is at risk of developing preeclampsia.

1.4 Research Questions

The main research question to be addressed by this research is:

RQ$_M$. *How can a machine learning model be used to support medical staff in identifying patients at risk of developing preeclampsia?*

The main research question is answered by addressing the following sub-research questions:

RQ 1. Is preeclampsia a significant risk factor for pregnant women?
RQ 2. What factors affect the chances of a woman developing preeclampsia during pregnancy?
RQ 3. What viable data source can be used to train the predictive model?
RQ 4. What existing techniques or methods can be used to predict preeclampsia?
RQ 5. Can a model be designed to accurately predict preeclampsia during pregnancy?
RQ 6. Can the model accurately predict preeclampsia?

1.5 Research Objectives

The main research objective of the research is:

RO$_M$. *Develop a machine learning model that can support doctors and midwives in identifying patients at risk of developing preeclampsia.*

The following sub-objectives were identified to achieve the main research objective:

RO 1. Investigate if preeclampsia is a significant risk for pregnant women.
RO 2. Identify the factors that may affect a woman's chances of developing preeclampsia during pregnancy.
RO 3. Identify a viable data source that can be used to train the predictive model.
RO 4. Identify existing techniques or methods that can be used to predict preeclampsia.
RO 5. Investigate if a model can be designed to accurately predict preeclampsia during pregnancy.
RO 6. Evaluate if the model can accurately predict preeclampsia.

1.6 Scope, Constraints and Ethics

This research aims to investigate if a model can be developed to predict if a pregnant woman is at risk of developing preeclampsia. A significant constraint that arose early in the study was the COVID-19 pandemic. COVID -19 affected everyone and every industry worldwide. Distancing and travel restrictions caused the majority of non-COVID-related studies to be suspended until further notice (Harper et al., 2020). These restrictions had a massive impact as they forced the study to consider alternative data sources, where the data is limited by what has been collected. Another aspect that needs consideration is the significant risk associated with having pregnant women who developed preeclampsia as part of the sample. Thus, following a strict and precise ethical procedure is crucial.

1.7 Research Methodology and Dissertation Structure

The Design Science Research (DSR) methodology was selected as an appropriate research methodology, as this project involves the design and development of an artefact (i.e., a model). DSR requires an in-depth literature review followed by a thorough comparison of existing systems (Peffers, Tuunanen, Rothenberger, &

Chatterjee, 2007). The information extracted is used to develop the artefact to contribute to the field of study (Hevner & Chatterjee, 2010). The DSR methodology comprises of the six phases illustrated in Figure 1-1.

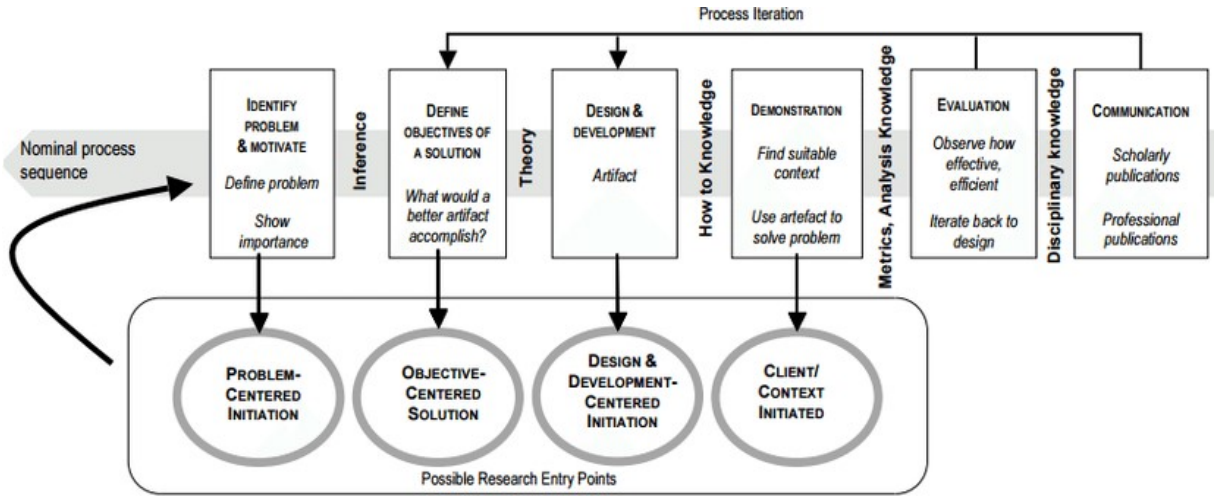

Figure 1-1 DSR methodology process map (Peffers, Tuunanen, & Gengler, 2006)

Chapter 1 - Introduction

Chapter 1 explains the background of pregnancy complications and their effects in Sub-Saharan Africa and, more specifically, South Africa. All the research criteria are also stated and explained.

Chapter 2 – Research Methodology

Chapter 2 included a motivation for the selection of DSRM, followed by an an explanation on how DSRM will be implemented in this study.

Chapter 3 – Preeclampsia and Its Associated Risk Factors

Chapter 3 starts with an in-depth review of relevant literature on pregnancy and some prominent causes of stillbirths, focusing on preeclampsia and its associated risk factors. The chapter will end with a review and comparison of existing models and systems.

Chapter 4 – Machine Learning in Healthcare

In the fourth chapter, a literature review is performed to identify the limitations and contributions of similar studies and models in healthcare. A discussion on machine learning and its impact on predictive models in healthcare is also provided.

Chapter 5 – Design and Development

In Chapter 5, the design and development process of the models is explained. This chapter also describes the development environment and tools used. An evaluation of the models and an in-depth analysis of the results are also included.

Chapter 6 – Conclusion

The dissertation's final chapter reflects on the data analysis and presents recommendations for future research, challenges and lessons learnt. The chapter ends with a summary of the dissertation and a few final thoughts.

Figure 1-2 illustrates the outline of the dissertation with respect to the DSRM activities.

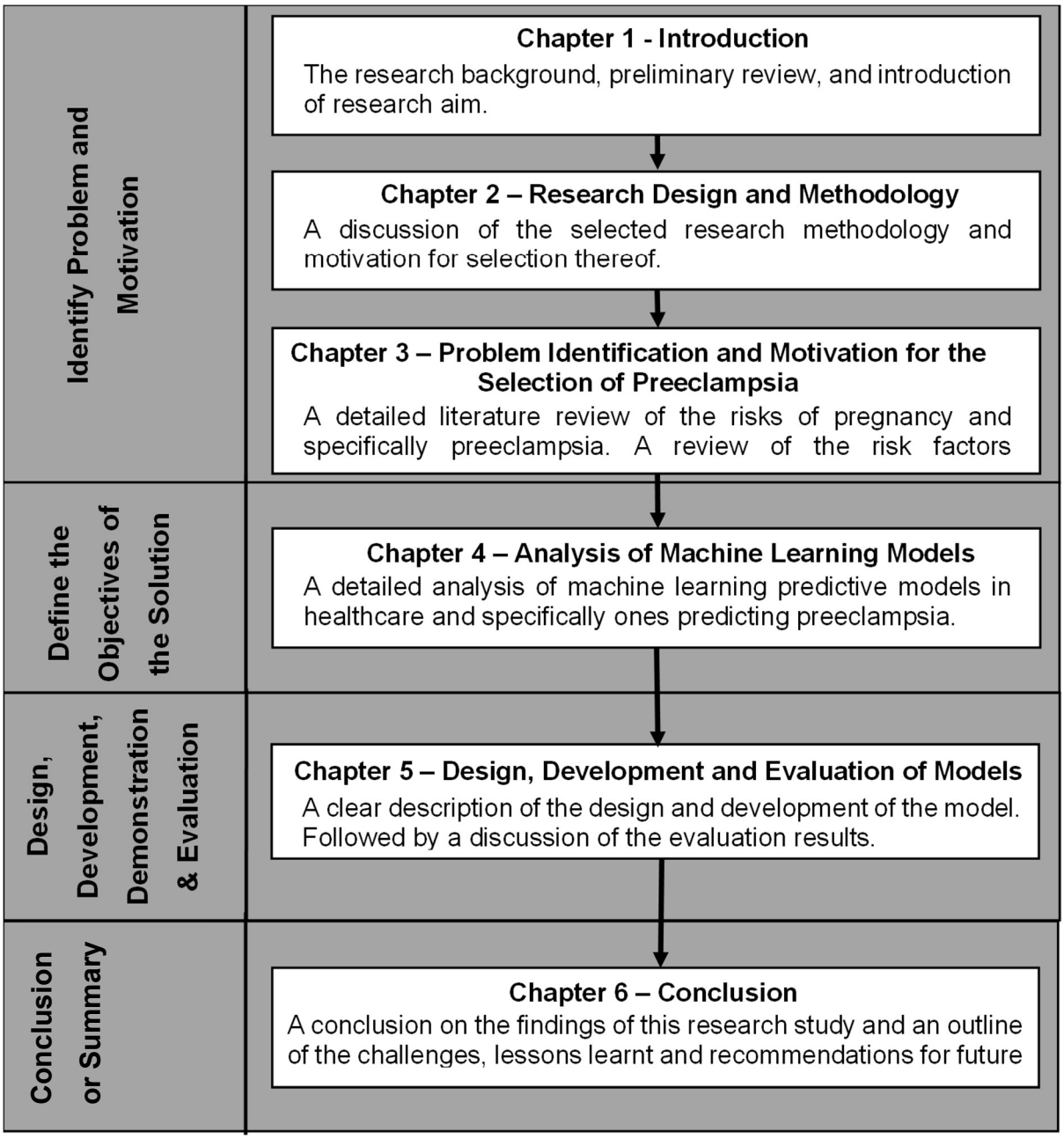

Figure 1-2 Dissertation Structure with Respects to the DSRM activities

Chapter 2. Research Methodology

2.1 Introduction

Chapter 1 introduced the problem and gave an overview of the dissertation structure. The Design Science Research Methodology (DSRM) is reviewed in Chapter 2, with an explanation of its context within this dissertation. A research design is a set of methods and procedures the researcher follows to achieve the research aim. A research design is also defined as a strategy adopted to incorporate all research components of a study in a logical format to address the research problem (Trochim, 2006).

This chapter is structured into nine sections to explain how DSRM will be implemented to answer the research questions. Section 2.2 includes a summary of the different research methodologies in Information Systems (IS), followed by a motivation for selecting DSR as this study's methodology in Section 2.3. The five DSR activities and an outline of the structure of this dissertation will be included in Sections 2.4 and 2.5, respectively. Section 2.6 compare the different evaluation types for DSR. Section 2.7 discusses how the application of the DSR methodology will guide this research. Ethical considerations applied in this study are detailed in Section 2.8, with a conclusion in Section 2.9.

2.2 Research Methodologies in Computer Science

Choosing the most suitable research methodology starts with defining whether the study is qualitative or quantitative (Table 2-1). A qualitative research methodology addresses questions that cannot be answered by quantification (Dixon-Woods, Agarwal, Jones, Young, & Sutton, 2005). Results drawn from qualitative methodologies should be descriptive, and the inferences should be easily drawn from the data. Quantitative research methods are a numeric or statistical approach to research design, and it creates meaning through objectivity uncovered in the data

collection (Williams, 2007). Quantitative research produces quantifiable, reliable, and verifiable data that can be generalised to a larger population.

Table 2-1 Comparison of Qualitative and Quantitative Research

Qualitative Research	Quantitative Research
Inductive approach	Deductive approach
Subjective Approach	Objective approach
Open and flexible	Closed and planned
Researcher is close to the respondents	Researcher is distant from the respondents
Theoretical sampling	Random sampling
Low-level measurement	High-level measurement

When looking at this study's research questions and objectives, it can be seen that this study follows a quantitative approach. However, a mixed methods approach could work well, and thus the DSRM would be a good research paradigm to follow. Section 4.3 motivates why DSRM is the selected research methodology.

2.3 Motivation for Selection of Design Science Research Methodology as the Research Methodology

The DSRM is the iterative development and evaluation of a solution to a real-life problem in artificial sciences (Hevner & Chatterjee, 2010). Researchers in IS use DSR to identify and investigate problems within the domain, followed by either improving or developing a solution (Baskerville, Baiyere, Gregor, Hevner, & Rossi, 2018). The key contributions required in DSR are:

- The design of a novel Information Technology (IT) artefact and the introduction of the artefact into an application context with measurable improvements (technology evolution).
- The addition of new prescriptive knowledge contributions in the form of IT artefacts and nascent design theories extends and generalises the knowledge contribution of the DSR project (technology informing science).

DSR is aimed at creating and evaluating a proposed artefact. This artefact should resolve the identified problem while satisfying the criteria for being a DSR knowledge contribution (Hevner, March, Park, & Ram, 2004). Two artefacts will be produced: a prediction model and a proof-of-concept prototype. Therefore, DSR is selected as the research methodology for this study.

2.4 DSR Activities

The DSRM process model (Figure 2-1) developed by Vaishnavi and Kuechler (2004) aims to assist researchers using the DSRM. Five activities in a nominal sequence form the model (Vaishnavi & Kuechler, 2004). These activities are explained in more detail below.

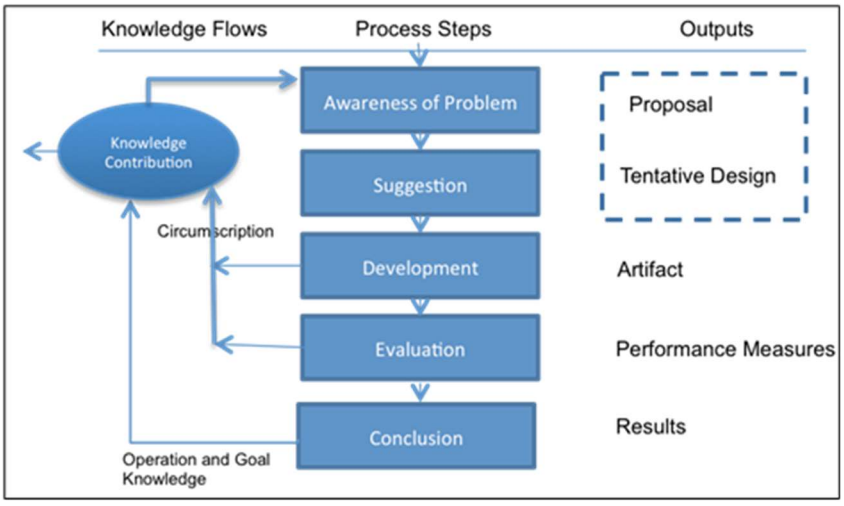

Figure 2-1 DSR Methodology Process Model **(Vaishnavi & Kuechler, 2004)**

[23]

Awareness of Problem

The first activity involves defining the research problem to be solved, followed by a motivation for why the proposed artefact is relevant and valuable (Peffers et al., 2006). It is crucial to understand the problem so that the proposed artefact can satisfy all identified aspects. Breaking the problem into smaller, more manageable parts will enable a better understanding of each aspect. This study's proposal, and Chapter 1, will cover all the details involved in identifying the problem. Defining the objectives of the solution is essential for determining its feasibility and motivating its relevance. These objectives must be clear and precise as they will be reference points and will guide the study as they advance. There are two types of objectives, namely, qualitative and quantitative. Quantitative objectives identify how the proposed model will be better than the existing ones. Qualitative objectives describe how the proposed model will support solutions to the problem. Chapters 1 and 2 are responsible for defining the research objectives.

Suggestions

Once the problem has been defined, the researcher should indicate the type of artefact to be developed. An artefact is an object devised by researchers to solve an existing problem in the contextual environment (Johannesson & Perjons, 2014). However, an artefact is not only tangible but can also be intangible such as a model, framework, or architecture. As highlighted previously, artefacts play an essential role in DSR. Johannesson and Perjons (2012) outline the different artefact types as Constructs, Descriptive Models, Prescriptive Models, Predictive Models, Methods, and Instantiations.

A model is a representation of one or more concepts aimed at solving a practical problem ("A Practical Guide. to SysML," 2015). There are four main types of models: descriptive, prescriptive, predictive, and explanatory models. The predictive model is

used for forecasting the behaviour of a system. Thus, the selected artefact type for this study is a predictive model.

Table 2-2 Outline of Different Artefact Types (Johannesson & Perjons, 2014)

	Description
Construct	The term, notation, definition, or concept that provides definitional knowledge. Aids in explaining practical problems for better comprehension thereof.
Descriptive Model	Describes and defines the existing practical problems in the contextual environment.
Prescriptive Model	Describes the potential solutions to the existing practical problems.
Predictive Model	Provides a forecast of the behaviour of systems or objects.
Method	Outlines prescriptive knowledge that defines guidelines and processes to solve the existing practical problems.
Instantiation	A working system or solution that is applied in the contextual environment to solve existing problems.

Design and Development

The design and development activity consist of designing and developing the proposed artefact based on the solution's requirements. The design involves understanding the domain and applying technical knowledge through a conceptual representation of the artefact. Development refers to the construction of the artefact from the design.

Evaluation

Evaluation of the artefact involves observing and measuring the extent to which the artefact solves the identified problem. Evaluation requires a thorough knowledge of relevant metrics and analysis techniques. The results from the evaluation will support the decision to perform an additional iteration or not. If an additional iteration is needed, the researcher can proceed to either the awareness, suggestion, or development phases.

Conclusion

The conclusion is the final phase, where the research results and contributions are identified and discussed. The following should be performed:

1. Discuss the problem and its relevance.
2. Present the artefact.
3. Discuss the contributions of the artefact.
4. Present the effectiveness of the artefact in solving the outlined problem.
5. Decide if the output of this phase is an acceptable research contribution.

2.5 Dissertation Structure

In an article by van der Merve, Gerber and Smuts (2019), they map the DSR cycle to the postgraduate research report. They aimed to analyse different types of postgraduate research reports that implemented DSR and develop a map between the structure of the report and Vaishnavi and Kuechler's (2004) DSR process model. Figure 4.2 illustrates how van der Merwe, Gerber and Smuts (2019) suggested mapping a master's dissertation with a single function artefact.

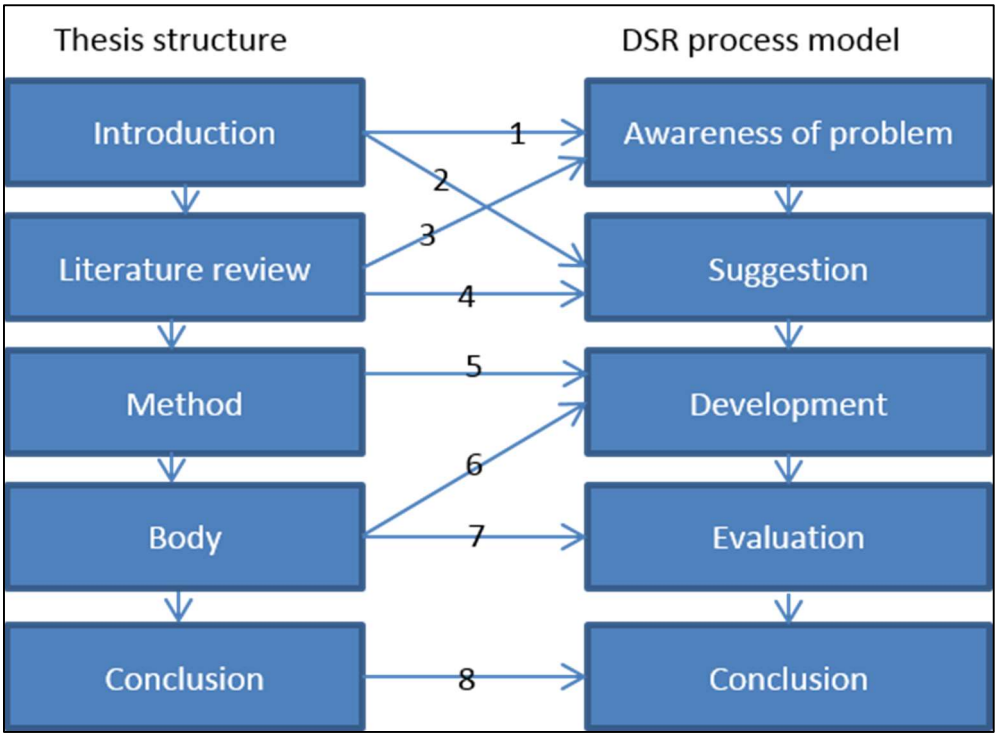

Figure 2-2 Map of Masters Dissertation Structure to DSR Process Model (Adapted from van der Merwe, Gerber, & Smuts, 2017)

Mapping 1 and 2: Introduction, awareness of the problem and suggestion. Chapter 1 introduces the problem and indicates the type of solution (artefact). Chapter 2 describes the strategy adopted to address the identified problem.

Mapping 3 and 4: Literature review, awareness of problem and suggestion. Performing a literature review allows for an extensive investigation of the literature. Chapters 3 and 4 will justify the selected problem by confirming that it is relevant, and has not previously been solved or how the proposed solution differs from existing solutions. A suggestion of an artefact is provided at the end of Chapter 4.

Mapping 5, 6 & 7: Method and Body, Development and Evaluation. Chapter 5 makes up the body of the dissertation and includes all data relevant to the building and evaluation of the artefact. A detailed description of the development of the artefact and

all experiments conducted is provided. The chapter concludes with the evaluation of the artefact and a discussion of the results.

Mapping 8: Conclusion. The final chapter of this dissertation, Chapter 6, will conclude the study and evaluate the research contribution.

Using the DSR research report map (Figure 2.1) to guide this dissertation structure allows for the successful implementation of the DSR methodology. In Section 2.6, the DSR process model activities are investigated.

2.6 Formative and Summative Evaluations in Design Science Research

The evaluation consists of examining the DSR outputs, including design artefacts and IS design theories (Jones & Gregor, 2007; March & Smith, 1995; Walls, Widmeyer, & El Sawy, 1992). Evaluation is critical as it provides evidence that the artefact developed in DSR achieves the aim of the study (Venable, Pries-Heje, & Baskerville, 2012). Without rigorous evaluation, DSR is an unsupported design theory where the developed artefact may not satisfy the aim. Unfortunately, DSR literature provides little guidance regarding the selection of evaluation strategies and methods. Venable et al. (2012) developed and presented an enhanced version of an existing DSR Evaluation Strategy Framework proposed by Pries-Heje, Baskerville, and Venable (2008). Venable, Pries-Heje, and Baskerville (2012) summarised five primary purposes for evaluation in the DSR literature:

- Evaluate an instantiation of a designed artefact to establish its utility and efficacy (or lack thereof) for achieving its stated purpose.
- Evaluate the formalized knowledge about a designed artefact's utility for achieving its purpose.
- Evaluate a designed artefact or formalized knowledge about it compared to other designed artefacts' ability to achieve a similar purpose.

- Evaluate a designed artefact or formalized knowledge about it for side effects or undesirable consequences of its use
- Evaluate a designed artefact formatively to identify weaknesses and areas of improvement for an artefact under development

Venable, Pries-Heje, and Baskerville (2012) extended Pries-Heje et al. (2008) in three parts: 1) a framework extension that allows the mapping of contextual aspects to a potential evaluation strategy or strategies, 2) a framework extension to map evaluation strategies to evaluation methods, and 3) a method for using the two framework extensions.

The first extension is a Strategy Selection Framework to assist the researcher in selecting under which strategy their study falls (Table 2-3). The researcher must identify their study's priority criteria and the appropriate quadrants. This study follows an artificial ex-post, possibly an ex-ante, evaluation strategy as the artefact for this study is purely technical, in the form of a model. However, Venable et al. (2012) noted that a hybrid strategy, combining more than one quadrant, is possible for studies which have conflicting goals.

Table 2-3 Venable et al.'s (2012) First Extension DSR Evaluation Method Selection Framework

DSR Evaluation Strategy Selection Framework		**Ex-Ante**	**Ex-Post**
		• Formative • Lower build cost • Faster • Evaluation design, partial prototype, or full prototype • Less risk to participants (during evaluation) • Higher risk of false positives	• Summative • Higher build cost • Slower • Evaluation instantiation • Higher risk to participants (during evaluation) • Lower risk of false positive
Naturalistic	• Many diverse stakeholders • Substantial conflict • Socio-technical artefacts • Higher cost • Longer time – slower • Organizational access needed • Artefact effectiveness evaluation • Desired Rigour: "Proof of the pudding" • Higher risk to participants • Lower risk of false positives	• Real users, real problems, and slightly unreal system • Low-medium cost • Medium Speed • Low risk to participant • Higher risk for false positive	• Real users, real problems, and real system • Higher cost • Higher risk to participant • Best evaluation of effectiveness • Identification of side effects • Lower risk of false positives
Artificial	• Few similar stakeholders • Little or no conflict • Purely technical artefact • Lower cost • Less time • Desired Rigor: Control of variables • Artefact efficiency evaluation • Higher risk of false positives	• Unreal users, problem, and/or system • Lowest cost • Fastest • Lower risk to participants • Highest risk of false positives (effectiveness)	• Real system, unreal problem, and possibly unreal users • Medium-high cost • Medium speed • Low-medium risk to participants

The second extension is the method selection framework, where the different evaluation strategies in Pries-Heje et al. (2008) are used to select different extent evaluation methods. As mentioned before, this study follows an artificial ex-post evaluation strategy. Table 2-4 maps the selected strategy to the appropriate evaluation methods.

Table 2-4 Venable et al.'s (2012) Second Extension DSR Evaluation Method Selection Framework

DSR Evaluation Method Selection Framework	Ex-Ante	Ex-Post
Naturalistic	• Action Research • Focus Group	• Action Research • Case Study • Focus Group • Participant Observation • Ethnography • Phenomenology • Survey (qualitative or quantitative)
Artificial	• Mathematical or Logical Proof • Criteria-Based Evaluation • Lab Experiment • Computer Simulation	• Mathematical or Logical Proof • Lab Experiment • Role Playing simulation • Computer Simulation • Field Experiment

The third extension is a four-step DSR evaluation research design method that relies on the extended framework. The four steps include:

1) An analysis of requirements for the designed evaluation,
2) The mapping of the requirements to one or more quadrants,
3) The use of Table 2-4 to select an appropriate evaluation method, and
4) The designing of highly detailed evaluations such as surveys or experiments.

Due to the proposed artefact being a machine learning model, steps 1,2 and 3 were answered in this section. The designing of the evaluation, step 4, is performed in the relevant chapter.

2.7 Application of DSR Guidelines

DSR has a set of guidelines used by researchers to understand, execute and evaluate their research. The following guidelines are applied to this study:

- Design as an Artefact – DSR studies must produce a viable artefact, and this study will produce a prediction model.
- Problem Relevance – The artefact should be technology-based and must be designed and developed within the study. This study will design and develop a model to assist healthcare workers in making more informed decisions when dealing with pregnant women with preeclampsia.
- Research contributions – Studies using DSR must prove that the study and accompanying artefact are valuable and helpful to the relevant fields. This study's contribution is a predictive model that can be applied by health workers working with pregnant women responsible for decision-making.
- Research Rigor – DSR requires rigorous design and evaluation of the artefact to ensure the quality of the solution to the outlined problem. This study applied iterative evaluations of the artefact to ensure research rigour.
- Design as a search process - The design of an artefact to solve existing problems in the contextual environment is an ongoing process, with each iteration aiming to improve the artefact further. In this research, the model went through several design and evaluation iterations.
- Communicate the research – The results of DSR need to be presented to the relevant audiences to allow sharing of knowledge within the field. This study will enable the author to publish papers in relevant conferences and journals.

2.8 Ethical Consideration

Myers and Venable (2014) proposed a set of ethical principles for DSR in IS. Participants are the people who are directly or indirectly involved in the study. They highlighted six basic principles a study should follow, acknowledging that each study will have a mix of these principles that apply to their study. The six principles are:

1) Public interest,
2) Informed consent,
3) Privacy, Honesty, and Accuracy,
4) Property, and
5) Quality of artefact.

Since this study used an existing dataset, the data supplier dealt with some of these principles. This study was responsible for the following principles.

- The public interest – Design science researchers must identify the stakeholders associated with their artefact. Generally, principles of safety, health, democracy, empowerment, and emancipation for all should predominate in choices of the features and capabilities that an artefact should have.
- Honesty and Accuracy – No plagiarism should happen in DSR, only inspiration from other sources. DSR also relies on the researcher's honesty in reporting on their findings, good or bad.
- Quality of artefact – As mentioned previously, research rigour is vital for the successful execution of DSR. As this study is high-risk, the design should account for and address such risks, and evaluation and testing must be sufficient to ensure safety in use.

All the data received from the supplier was anonymised, ensuring patients' privacy was not impeded. Finally, all data is owned by the data supplier, and the researcher followed all protocols that the data supplier required. A medical dataset was used,

which required ethical clearance from the Ethics department at Nelson Mandela University. Appendix A shows the ethical clearance approval letter provided by NMU.

2.9 Conclusions

This chapter reported on which research methodology suits this study and described how it was implemented. Vaishnavi & Kuechler's (2004) five main objectives were discussed and mapped to this dissertation chapters using van der Merwe, Gerber, and Smuts' (2019) dissertation map. Thereafter, an examination of the different evaluation techniques was included. Venable et al. (2012) evaluation framework was selected and identified five evaluation methods that could guide this study. These methods included:

- Mathematical or Logical Proof
- Lab Experiment
- Role Playing simulation
- Computer Simulation
- Field Experiment

This chapter supports the selection of DSR as the chosen methodology. The next chapter includes the next phase of research, which is reviewing the literature and comparing existing models

Chapter 3. Literature Review

3.1 Introduction

This chapter conducted a literature review on pregnancy and some prominent causes of stillbirths. This chapter aims to answer the following research questions:

RQ 1. Is preeclampsia a significant risk for pregnant women?

RQ 2. What factors affect the chances of a woman developing preeclampsia during pregnancy?

The chapter is structured into five sections and starts with a description of the literature review process. Section 3.3 contains a brief discussion on pregnancy followed by a literature search to identify risk factors to support the claim that hypertension is one of the most prevalent risk factors. In Section 3.4, a more thorough literature review is conducted on hypertension, specifically preeclampsia, during pregnancy, followed by an investigation into existing systems and models in Section 3.5. Finally, Section 3.6 concludes what has been done in the chapter and what will be covered in the next chapter.

3.2 Literature Review Process

The researcher used the following databases to find relevant research: Google Scholar, ResearchGate, Springer, IEEE Xplore, ScienceDirect, Mendeley, and PubMed. The inclusion search criteria were the following:

- Focus on pregnancy, and its associated conditions/diseases that may result in a stillbirth,
- Focus on the factors leading to hypertension/preeclampsia during the pregnancy period, and
- Be published in English.

The review aimed to indicate how this study fits in with what other scholars have found on preventable risks for women during pregnancy. The following sections report on the review's findings according to the main topics and themes.

3.3 Pregnancy and its Associated Risk Factors

Pregnancy is a complex period involving many stages, leaving room for many risks. This section focuses on identifying risk factors that may lead to the adverse outcome of a stillbirth. A risk factor is something that increases a patient's risk of developing an adverse outcome (Offord & Kraemer, 2000). WHO defines stillbirth as the death of a foetus late in pregnancy and requires countries to define at what week a miscarriage becomes a stillbirth (World Health Organisation, 2016). In South Africa, the Births and Deaths Registration Act defines stillbirth as an infant born 26 weeks after conception (Births and Deaths Registration Act 51 of 1992, 1992). In 2015 approximately 2.6 million babies were stillborn (De Bernis et al., 2016). Usually, fully functioning maternity services should detect complications by the time labour starts and provide the mother and families with all the necessary interventions, quality maternal advice and care for a healthy new-born. Unfortunately, this is not the case all over the world. In LMICs, maternal care is a healthcare facility that is not prioritised.

First, a systematic literature review by Aminu et al. (2014) was consulted. Aminu et al. investigated 142 studies, of which 49 were from Africa, retrieved from electronic databases such as MEDLINE, CINAHL Plus, Global Health, and LILACS for studies on stillbirths or disease conditions leading to stillbirths in LMICs between 2000 and 2013. Aminu et al. (2014) included studies that met the following criteria:

- Must have assessed at least one of the causes or risk factors of stillbirths (irrespective of the definition of stillbirth used),
- Were conducted in LMICs as defined by the World Bank, and
- Were published in English.

Table 3-1 Summary of Aminu et al.'s (2014) findings

Attributed cause of stillbirth	% cause range	No. of studies reporting causes	Total no. of stillbirths reported on	No. of stillbirths per study
Mother's disease: e.g., diabetes, Human Immunodeficiency Virus (HIV), syphilis	8–50	21	6392	12–1748
Foetal: e.g., congenital anomalies, infections	2.1–33.3	16	3040	12–640
Placental: e.g., placenta praevia, placental abruption	7.5–42	12	3024	12–640
Intrapartum: e.g., asphyxia, birth trauma	3.1–25	6	1094	24–735
Umbilical: e.g., prolapse, loop, knot	2.9–12	6	660	17–266
Trauma: e.g., iatrogenic	5–28	3	901	32–735
Amniotic: e.g., chorioamnionitis, oligohydramnios	6.5	1	169	169
Uterine: e.g., rupture, anomalies	10.7	1	169	169
Unclassified/unknown/unexplained	3.8–57.4	16	5313	12–1748

[37]

Maternal disease was the most prevalent risk factor, with 21 studies identifying one or more maternal diseases as the cause of stillbirth (Table 3-1). These maternal diseases included hypertension and preeclampsia. An important suggestion given by Aminu et al. (2014) was that a reduction in stillbirths could be achieved by improving the uptake and quality of intrapartum care for women. However, they acknowledged the fact that [34] it varies from country to country. Aminu et al.'s (2014) study highlights the need for better intrapartum care to lower the number of maternal diseases leading to stillbirths. The main problem in LMICs is the availability of medical professionals that can provide quality intrapartum care. There is less than one doctor available for every 1000 people in South Africa. This statistic is even worse in some provinces, such as Limpopo, where there are only 0.2 per 1000 people (Ayo-Yusuf, 2015; Wildschut, 2010). Brazil, a country with a similar economic situation, has nearly double this. If this wasn't bad enough, there are only approximately 3.2 nurses or midwives for every 1000 people. The lack of medical professionals has been an ongoing problem in South Africa. In an interview conducted on 09 May 2022, health minister Phaala reported that currently, we had 0.31 doctors per 1000 patients. A decrease from 2019, when South Africa had 0.79 doctors per 1000 patients (Clarke, 2022).

A population-based study by Gardosi, Madurasinghe, Williams, Malik, and Francis (2013) assessed the main risk factors associated with stillbirths in a multi-ethnic English maternity population. The study was conducted in England's National Health Service (NHS) regions. They used the NHSnet database's perinatal episode electronic records (PEERs), which are hosted and managed by the West Midlands Perinatal Institute. They made use of data collected between June 2009 and May 2011. After an initial exploratory analysis, Poisson regression models were used to assess explanatory variables' independent and multiple variable effects on stillbirths. The variables they used were those known to have clinical relevance and from previous publications. Factors were split into six groups: general maternal characteristics, social

factors, maternal history, pregnancy-related factors, complications in pregnancy, and foetal/neonatal characteristics. A significant risk factor identified was a parity of three or higher, which increased a mother's risk by 60 per cent. Contrary to the systematic review consulted, 77 per cent of studies found a statistical significance between maternal age and stillbirths, whereas Gardosi et al. (2013) found no significance. They also found that mothers living in the most disadvantaged areas had an increased risk of pre-existing conditions such as hypertension. Diabetes, and a history of mental health problems, also had a higher risk. Body Mass Index (BMI) categories of 30-34.9 and 35 or more represented a 40 per cent and 60 per cent increase in the risk of stillbirth, respectively. The highest stillbirth rate was among non-smoking pregnant women with foetal growth restriction. Gardosi et al. (2013) attribute this to these pregnancies being considered low-risk, and foetal growth restriction is less likely to be detected antenatally. Another interesting finding was that among the cohort of 389 stillbirths, the detection rate was very low. Of the cases, 195 (50.1%) had foetal growth restriction, and in 160 cases (82.1%), foetal growth restriction was not detected antenatally. Madhi et al.'s (2019) paper "Causes of stillbirths among women from South Africa: a prospective, observational study" shows a more detailed breakdown of the risks among South African women. For 13 months, Madhi et al. studied 354 stillbirths. Their research was conducted in Soweto at the Chris Hani-Baragwanath Hospital, where they looked at stillbirths of foetuses of at least 22 weeks' gestational age or with a birth weight of at least 500g. Of the 354 stillbirths enrolled, complete samples were available for 298 [36] stillbirths born to 294 mothers. In Appendix B, the demographic and baseline clinical features of women who had stillbirths can be seen. Among the 289 complete samples, 243 were diagnosed predelivery by ultrasonography, and only 46 were diagnosed at delivery. The most prevalent maternal medical condition that resulted in a stillbirth was hypertension, with 56 cases (Table 3-2). The closest second was diabetes, with only 6 cases. The prevalence of hypertension in Madhi et al. (2019) supports the claim that hypertensive disorders are a significant risk for pregnant women.

***Table 3-2 Main maternal or foetal conditions that possibly or probably contributed to the stillbirth* (Madhi et al., 2019)**

Maternal or Foetal Condition	Total (n=298)	p value*
Maternal medical condition during pregnancy	64 (21%)	0·14
Hypertensive disorder	56 (19%)	..
Diabetes	6 (2%)	..
Other	2 (1%)	..
Clinical obstetric complications	54 (18%)	0·53
Clinical chorioamnionitis	1 (<1%)	..
Intrapartum foetal distress with asphyxia or hypoxic intrapartum foetal distress	2 (1%)	..
Placental abruption	45 (15%)	..
Uterine rupture	5 (2%)	..
Uteroplacental insufficiency	1 (<1%)	..
Foetal, genetic, or structural abnormality	6 (2%)	0·42
Infection	58 (19%)	>0·99
Placental infection and decreased placental function	11 (4%)	..
Foetal bacterial infection	47 (16%)	..
Pathological placental conditions	57 (19%)	0·88
Placental disc	1 (<1%)	..
Placental membranes	1 (<1%)	..
Umbilical cord	1 (<1%)	..
Foetal membrane and placental inflammation	27 (9%)	..
Circulatory abnormalities	26 (9%)	..

Other placenta abnormalities	1 (<1%)	..
Other	4 (1%)	0·62
Unknown	55 (18%)	0·22

Battarbee, Sinkey, Harper, Oparil, and Tita (2020) reviewed the challenges associated with hypertension among pregnant women in the United States of America (USA). They classified hypertension into two main groups: chronic hypertension and pregnancy-induced hypertension. Chronic hypertension is one of the most common medical disorders and is defined as hypertension diagnosed before pregnancy or before 20 weeks' gestation. After 20 weeks' gestation, it is defined as pregnancy-induced hypertension. Madhi et al.'s multicentre study used a model to assign the cause of death using clinical registry data. The model identified intrauterine asphyxia as the most prevalent cause of stillbirths, followed by underlying maternal conditions such as prolonged labour and preeclampsia.

In the final study reviewed, a panel of top researchers in the field reviewed potential risk factors for stillbirths, including 38 maternal factors (Lawn et al., 2016). Their goal was to review the literature to identify risk factors associated with stillbirth. They included risk factors that had a high correlation to stillbirths and available prevalence data for all countries worldwide. Systematic literature reviews published between 2010 and 2015 were analysed to identify any data that could infer the risk association. Lawn et al. (2016) highlights the importance of preventing risk factors such as infections during pregnancy. For example, malaria is estimated to be attributable to about 20 per cent of stillbirths in sub-Saharan Africa. Improvements in preventing and treating malaria in pregnancy should be an essential first step to preventing stillbirths in countries with weak health systems. Lawn et al. (2016) also mention that more than 200 000 stillbirths are attributable to preeclampsia and eclampsia, with sub-Saharan Africa and South Asia being the most impacted regions. These deaths could be

prevented with detection and appropriate management of risk factors during antenatal care. Lawn et al. (2016) acknowledges that many LMICs have poor antenatal care. An increase in the quality of antenatal care could identify and address many of these disorders.

After consulting these eight studies by experts in the field, it is clear that hypertension associated with pregnancy is a significant risk factor that increases the likelihood of stillbirth. These findings support this research's aim and show how important it is to identify hypertension during pregnancy. Section 3.4 provides a deeper understanding of what hypertension is and the risk factors associated with it.

3.4 Hypertension Associated with Pregnancy

As mentioned in Section 2.2, hypertension during pregnancy can be classified into two categories: Chronic hypertension and hypertensive disorders of pregnancy (HDPs). Chronic hypertension, which needs to be diagnosed before gestation, was not considered as it was out of this project's scope. Therefore, only HDPs were investigated further.

HDPs are among the most prevalent risk factors affecting approximately 10 per cent of pregnancies. Pregnant women with HDPs have an increased risk of long-term hypertension, stroke, cardiovascular mortality and major adverse cardiovascular events (Ananth, Keyes, & Wapner, 2013b). Sixteen per cent of maternal deaths in developed countries were attributed to hypertensive disorders (Khan, Wojdyla, Say, Gülmezoglu, & Van Look, 2006). The foetus is also at risk of developing intra-uterine growth restriction, preterm birth, placental abruptions, foetal distress, and foetal death (Haddad et al., 2004; Madazli et al., 2014; Rezk, Gamal, & Emara, 2015). Currently, the only treatment for HDPs is Aspirin, which is recommended in low dosages from the 12th week to delivery (Akbari, Khodadadi, Ahmadi, Abbaszadeh, & Shahsavar, 2018). Another method medical professionals recommend to reduce preeclampsia is lifestyle interventions such as adopting a higher plant-based diet (Brantsæter et al.,

2009; North et al., 2011). HDPs can be split into two categories: 1) Gestational Hypertension and 2) Preeclampsia.

3.4.1 Gestational Hypertension

Gestational hypertension (GH), also known as pregnancy-induced hypertension is defined as new hypertension in pregnant women after 20 weeks of gestation without the presence of protein in the urine or other signs of preeclampsia (Chobanian et al., 2003). Hypertension during pregnancy is a significant health issue for women and their babies worldwide. GH affects roughly five per cent of pregnancies (Garovic & Hayman, 2007; “Hypertens. Pregnancy,” 2013). In a study by Shen et al. (2017), they identified significant risk factors for gestational hypertension. They found that risk factors include a BMI of less than 25, nulliparity, preeclampsia history, type 1 diabetes, type 2 diabetes, and twin births. These can be split into two categories, history or static factors and dynamic factors. Parity and preeclampsia history are two historic/static factors that must be obtained before using the proposed model. The final three: BMI, diabetes and blood pressure, are dynamic factors since they have the potential to change. These three factors can be monitored at multiple different stages.

3.4.2 Preeclampsia

Preeclampsia is a pregnancy disorder affecting approximately 4.6 per cent of all pregnancies (Abalos, Cuesta, Grosso, Chou, & Say, 2013). It remains a leading cause of maternal and perinatal morbidity and mortality worldwide (Rosser & Katz, 2013; Savaj & Vaziri, 2012). In guidelines published by the National Institute for Health and Care Excellence in 2019, they classify a woman as at high risk of preeclampsia if there is the presence of one of the following risk factors:

- History of hypertensive disease during a previous pregnancy,
- History of maternal disease, or
- Diagnosed with chronic hypertension.

Women are at moderate risk if they have the presence of two or more of the following risk factors:

- Nulliparous,
- Are older than 40 years of age,
- A BMI ≥ 35 kg/m,
- A family history of preeclampsia,
- A multifetal pregnancy, or
- A pregnancy interval of more than ten years.

The presence of one high-risk factor, or two or more moderate risk factors, is used to help guide the prescription of aspirin prophylaxis (Ananth, Keyes, & Wapner, 2013a; M. A. Brown et al., 2018; NICE, 2019). If aspirin prophylaxis is administered before 16 weeks of pregnancy, it can drastically reduce the risk of developing preeclampsia (Askie, Duley, Henderson-Smart, & Stewart, 2007; Bujold et al., 2010). These risk factors are supported in the largest meta-analysis of clinical risk factors conducted by Bartsch et al.'s (2016) supported the significance of these risk factors by conducting a large meta-analysis of risk factors, analysing over 25 million pregnancies from 92 studies.

Blood Pressure

Blood pressure is one of the two leading indicators of preeclampsia and is measured with a sphygmomanometer. Mothers usually have their blood pressure measured during the pregnancy at each antenatal check-up. It is recommended that if a mother is at risk of preeclampsia, she monitors her blood pressure more frequently than just at antenatal check-ups (M. A. Brown et al., 2018). A blood pressure of 140 mmHg systolic or a diastolic pressure of 90 mmHg measured twice to confirm true hypertension is part of the diagnostic criteria for preeclampsia.

Proteinuria

Proteinuria, the presence of protein in the urine, is considered abnormal if it is greater than 300 mg in 24 hours (Airoldi & Weinstein, 2007). It is measured using a urine dipstick. Proteinuria is typically only measured when the gynaecologist, obstetrician, or midwife suspects the patient is at risk of developing preeclampsia. Since the model was designed to predict preeclampsia before a proteinuria test would occur, there was no need to include it as a risk factor.

Personal/Family History of Hypertension Disorders

If the mother, or someone in her immediate family, has a history of hypertensive disorders, it can significantly increase the risk during the pregnancy. It was found that a family history of hypertensive conditions was a significant risk factor. Furthermore, Bezerra et al. (2010) found that if a pregnant woman's mother or sister, more specifically both, had a history of either preeclampsia or hypertension, it put the pregnant woman at higher risk.

Chronic hypertension is when the blood pressure is above 140/90 mmHg before conception, or 20 weeks of gestation. Severe chronic hypertension is then classified by a diastolic reading of 110 mmHg or higher. Women with chronic hypertension are at a higher risk of developing superimposed preeclampsia (Lindheimer, Taler, & Cunningham, 2008). Superimposed preeclampsia refers to a woman with chronic hypertension developing preeclampsia.

Disease

Mother and foetus diseases are a significant risk factor during pregnancy and can increase the risk of developing preeclampsia. Mothers and their foetus' can both develop diseases during the pregnancy, or the mother could have chronic diseases before the pregnancy that they will take into the pregnancy. These diseases include diabetes, congenital anomalies, infections, HIV, and syphilis.

Nulliparous

Nulliparous refers to women who have not yet borne a child. If a woman has had one or many stillbirth/miscarriages, but did not have a live delivery, they are still considered to be nulliparous. Nulliparous women are at a higher risk of developing preeclampsia than women who have had a successful pregnancy and birth (North et al., 2011).

Age

Advanced maternal age is a significant risk factor for many pregnancy complications, including preeclampsia. Many studies have observed that women of advanced age have a higher risk of developing preeclampsia, approximately double in some studies (Lamminpää, Vehviläinen-Julkunen, Gissler, & Heinonen, 2012).

Body Mass Index

BMI is a statistical index based on weight and height measurements for estimating body fat in people of any age (Weir & Jan, 2021). The following defines the different categories of BMI:

- Severely underweight - BMI less than 16.5kg/m^2
- Underweight - BMI under 18.5 kg/m^2
- Normal weight - BMI greater than or equal to 18.5 to 24.9 kg/m^2
- Overweight – BMI greater than or equal to 25 to 29.9 kg/m^2
- Obesity – BMI greater than or equal to 30 kg/m^2
    - Obesity class I – BMI 30 to 34.9 kg/m^2
    - Obesity class II – BMI 35 to 39.9 kg/m^2
    - Obesity class III – BMI greater than or equal to 40 kg/m^2 (also referred to as severe, extreme, or massive obesity)

WHO reported that the prevalence of overweight and obese women in the USA and South Africa is approximately 77 and 69 per cent, respectively (P. Brown, 2018). Many studies have found a strong correlation between high BMI and preeclampsia. A 2008 study found the adjusted risk of developing preeclampsia doubled for overweight women, and tripled for obese women (Hauger, Gibbons, Vik, & Belizán, 2008).

3.5 Existing Models and Systems

Repeated BP measurements and symptom reporting are powerful resources that can supplement decision-making in pregnancy care when integrated into a single platform (van den Heuvel et al., 2019). However, little research has been conducted on such a platform. van den Heuvel et al. (2019) developed SAFE@HOME, a platform that integrates a preeclampsia symptom checklist with a BP monitor. Fourteen women were monitored with the iHealth Track application to obtain correct measurements. An automated non-invasive oscillometer device was used in the study, which they validated for use in pregnancy. The oscillometer devices automatically transfer the data to the mobile application through Bluetooth. Along with the oscillometer, the mobile application collects data on symptoms via a Yes/No questionnaire. The questionnaire also included general pregnancy symptoms, which mothers should continually pay attention to throughout their pregnancy (Table 3-3).

Table 3-3 SAFE@HOME questionnaire (van den Heuvel et al., 2019)

QUESTIONS
Do you have a headache?
Do you have visual problems?
Do you have a tight, band-like feeling around the upper stomach?
Do you experience severe upper abdominal pain?
Do your fingers feel numb?
Do you feel nauseous?
Do you have ankle, hand or face swelling?
Do you have contractions?
Do you have vaginal fluid loss?
Do you have vaginal bleeding?

Participants submitted their symptom checklist along with their BPs for three consecutive weeks from Monday to Friday before 10:00 AM. They set a threshold of a systolic value of greater than 140 mmHg, or a diastolic of less than 90 mmHg, or an increase of 20 mmHg compared to the previous measurement (Tranquilli et al., 2014; Tucker et al., 2017, 2018). van den Heuvel et al. (2019) achieved a compliance rate of approximately 85 per cent. The alert system was accurate, and the questionnaire proved to be of additional clinical value.

Table 3-4 Blood Pressure Alarm Summary (van den Heuvel et al., 2019)

	ALARM TYPE	N (%)	AFTER MANUAL CHECK	CLINICAL IMPACT
BP ALARMS	Exceeded threshold	4 (2.1 %)	No False Positives	Diagnosis of chronic hypertension in one participant
	20 mmHg raise	2 (1.7 %)	No False Positives	Because of absence of preeclampsia symptoms; expectant management
NO BP ALARM		179 (96.2)	No False Positives	
TOTAL BP SUBMITTED		186		

van den Heuvel et al. (2019) found that although an alert for symptoms would be triggered, after reviewing the BP readings, there was no need for further action. Another finding was the pregnant women's increased willingness to perform repeated self-measurement compared to the average person. Over 98 per cent of women with hypertension reported positive feedback on their involvement. van den Heuvel et al. (2019) noted that although the thresholds were adjustable for each participant, it would require the doctors to create accounts for the participants. Many doctors may be reluctant to create and manage separate accounts for each patient. If the proposed model is implemented in a clinical setting, or something similar, an alternative approach for patient enrolment will need to be investigated. van den Heuvel et al.'s

(2019) study shows the importance of static and dynamic data for monitoring and predicting preeclampsia.

In a recent study published by Stanford University, Li et al. (2022) looked at improving risk prediction for preeclampsia by modelling pregnancy trajectory. The data source was acquired via a digital phenotyping algorithm that curated a dataset of 108,557 pregnancies from Electronic Health Records (EHRs) across the Mount Sinai Health System in the USA. Their first goal was to use the EHRs to identify novel features associated with preeclampsia risk. Features are independent variables used in machine learning and make up the dataset used as an input for the model. The measurable quantities of identified risk factors are referred to as features when referring to them in the context of a machine learning model. Li et al. (2022) identified patients who developed preeclampsia using a digital phenotyping algorithm. The algorithm identified 10 per cent of the population who developed preeclampsia, which is consistent with the literature's 2-10 per cent. They divided their dataset into 19-time points and used different feature selection methods for each (Figure 3-1). Seventyeight, 68, and 48 features were identified across the ante-, intra-, and postpartum periods, respectively, from their developed network of features. Out of these identified features, 21 were shared among all three periods. Furthermore, half of them were supported in the literature as being associated with preeclampsia risk. These features included: maternal age, systolic blood pressure (SBP), diastolic blood pressure (DBP), weight, gestational hypertension, haemoglobin, white blood cell count, chronic hypertension, preeclampsia history, and headaches.

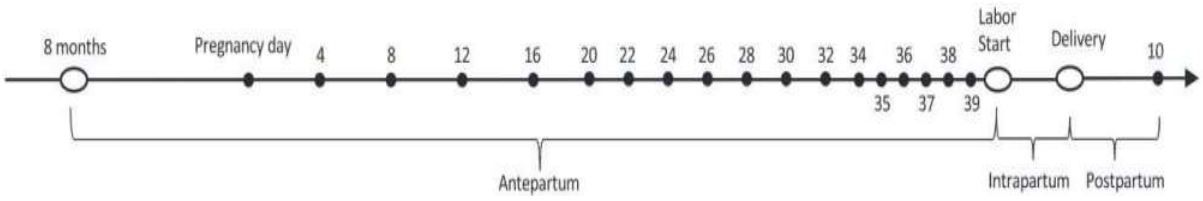

Figure 3-1 Li et al.'s (2022) dataset structure consisting of 19-time points

Due to the complex non-linear interactions among the extracted features, Li et al. used gradient boosted tree models. The reason for using gradient boosted trees are the model's ability to address missing values inherently. Subsequently, clinical features can be retrieved to avoid the bias and variation induced by imputation by standard methods such as mean, median, etc. Shapley values were implemented using the SHAP Python package to obtain global and local interpretability (Figure 3-2). Shapley values were introduced in 1951 by Lloyd Shapley as a solution concept in cooperative game theory. They can be used in many scenarios when the contribution of features that work cooperatively is unequal. These values aim to assign gains and costs to all the variables equally. Figure 3-2 supports the significance of features previously identified in the literature. Features that impacted the model include gestational hypertension, headache, maternal age, SBP, DBP, and twin pregnancies.

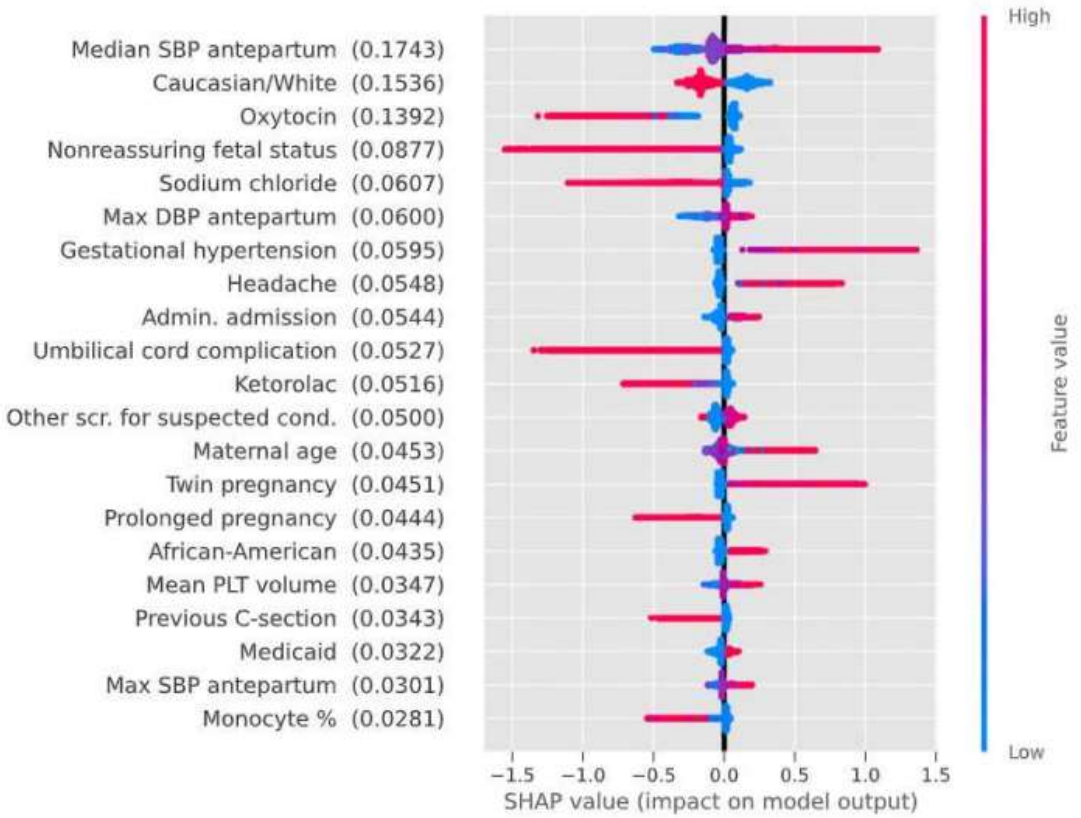

Figure 3-2 Shapley values used to identify the contribution of features

Li et al. (2022) found that interesting association patterns were observed by generating the moving average plots for the significant risk factors. One example of this appeared when examining SBP across the entire dataset. The data revealed that patients who developed preeclampsia in the antepartum period had higher SBP than those without preeclampsia. Li et al. (2022) highlighted the importance of using data from multiple time points to understand the data better. Multiple-time points can be extracted by aggregating live data over a period or by collecting data at specific time points. Deciding between these two methods comes down to the data collection technique used. Li et al.'s (2022) study emphasises the risk factors identified throughout the literature review.

3.6 Conclusions

The aim of Chapter 3 was to answer research questions one and two:

RQ 1. Is preeclampsia a significant risk for pregnant women?

RQ 2. What factors affect the chances of a woman developing preeclampsia during pregnancy?

Sections 3.3 & 3.4 highlighted the prevalence of preeclampsia as a significant risk factor during pregnancy, answering RQ 1. Sections 3.4 and 3.5 outlined an initial list of risk factors associated with developing preeclampsia, answering RQ 2. Multiple studies were reviewed to get a comprehensive set of risk factors, as follows:

- DBP,
- SBP,
- Maternal age,
- BMI,
- Diabetes status,
- Hypertension History,
- Nulliparity, and

- Results of a Health Questionnaire.

In the fourth chapter, a literature review is performed to identify the limitations and contributions of similar studies and models in healthcare. A discussion on machine learning and its impact on predictive models in healthcare is also provided.

Chapter 4. Machine Learning in Healthcare

4.1 Introduction

In this chapter, a literature review of predictive models in healthcare is provided. This chapter aims to answer the following research questions:

RQ 3. What viable data source can be used to train the predictive model?

RQ 4. What existing techniques or methods can be used to predict preeclampsia?

In Section 4.2, a discussion, supported by literature, is given on machine learning and the different aspects of development. An analysis of existing statistical and machine learning predictive models in healthcare is included in Section 4.3. Section 4.4 concludes the chapter by answering RQs 3 and 4.

4.2 Predictive Models in Healthcare

The healthcare sector is continually striving to achieve the Triple Aim, which improves outcomes, enhances patients' experience, and reduces healthcare costs to the public (Berwick, Nolan, & Whittington, 2017). Predictive modelling for real-time decision-making plays an essential part in achieving the Triple Aim. A clinical risk prediction model is defined as a model that combines several characteristics to predict the risk of disease, or the condition's presence and outcome occurrence in individuals (Alonzo, 2009). Risk prediction models, as decision-making tools, have long played an essential role in clinical practice. However, as times have changed, so do risk prediction models. Two significant changes in risk prediction have been centred around advances in Computer Science. Historical datasets are being replaced with live data streamed via sensors, and statistical models replaced with complex machine learning algorithms.

Dr Geoffrey Hinton, a well-cited researcher in machine learning, published a paper entitled "*Deep Learning – A Technology with the Potential to Transform Healthcare*".

He describes the importance of machine learning and, more specifically, deep learning in healthcare (Hinton, 2018). Deep learning is a sub-section of representation learning, which is a sub-section of machine learning (Figure 4-1).

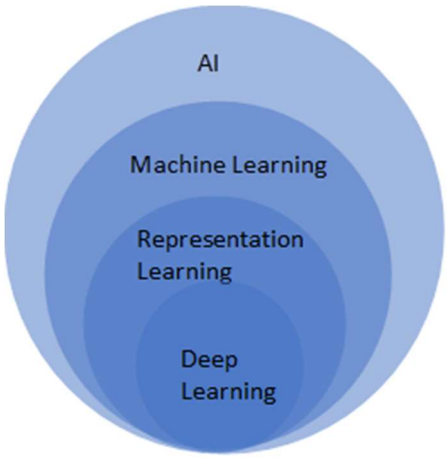

Figure 4-1 Representation of Sub-sections of AI **(Hinton, 2018)**

Deep learning allows multiple processing layers of a computational model to learn data representations with multiple levels of abstraction (Lecun, Bengio, & Hinton, 2015). The main difference between deep learning and machine learning is the amount of human intervention needed. Machine learning requires human intervention in some way, and commonly follows an iterative development process until a satisfactory prediction can be made. In contrast, deep learning models use a complex and intertwined neural network to identify if it was successful at predicting the outcome without any human intervention (Figure 4-2).

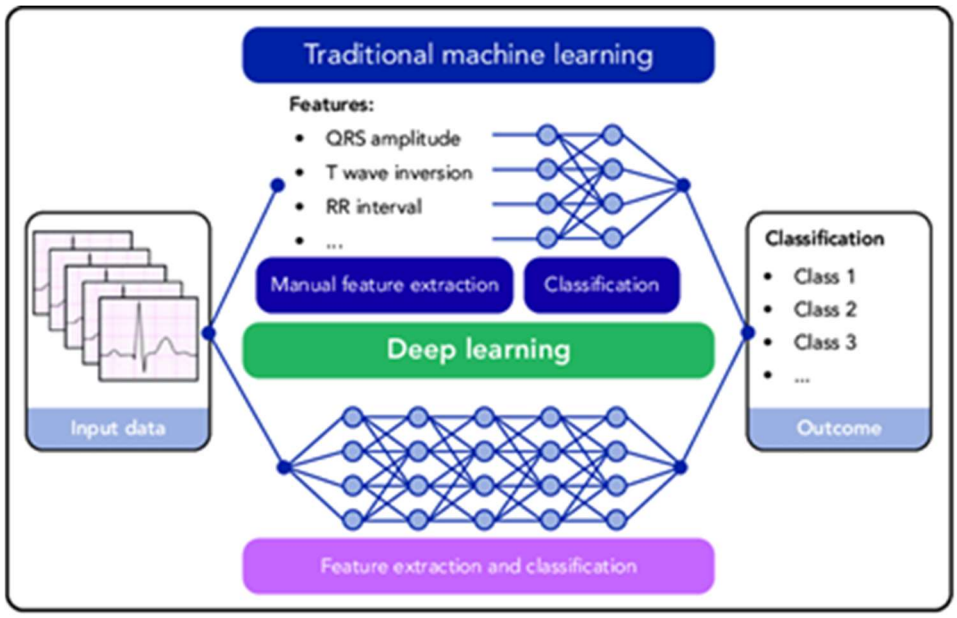

Figure 4-2 Representation of Deep Learning vs Machine Learning **(van de Leur et al., 2020)**

Dr Hinton (2018) suggests the use of deep learning in healthcare due to the number of complex interactions that need to be modelled. In many machine learning models, the developer will hand-select each feature based on knowledge and literature, whereas a deep learning model learns feature detectors optimised for classification. These features are very sensitive and require scaling to allow the model to assign a fair weighting to them. A scaler function is used to change the range of the data from a feature without changing the distribution. If not scaled, features may appear more significant than they actually are. An example of this is when assigning importance to a feature with high values such as the cost of purchase of a house compared to the number of cars in the garage; the much larger average value may cause the model to assign a higher importance to the feature. A significant disadvantage of using deep learning models is their required amount of data. Large field studies or clinical trials can become very complicated and high risk when conducting research in the health field. The likelihood of this project acquiring a data set large enough for a deep learning model is unlikely. Figure 4-3 shows an adaptation

of Raschka & Mirjalili (2019)'s roadmap for the development of machine learning models.

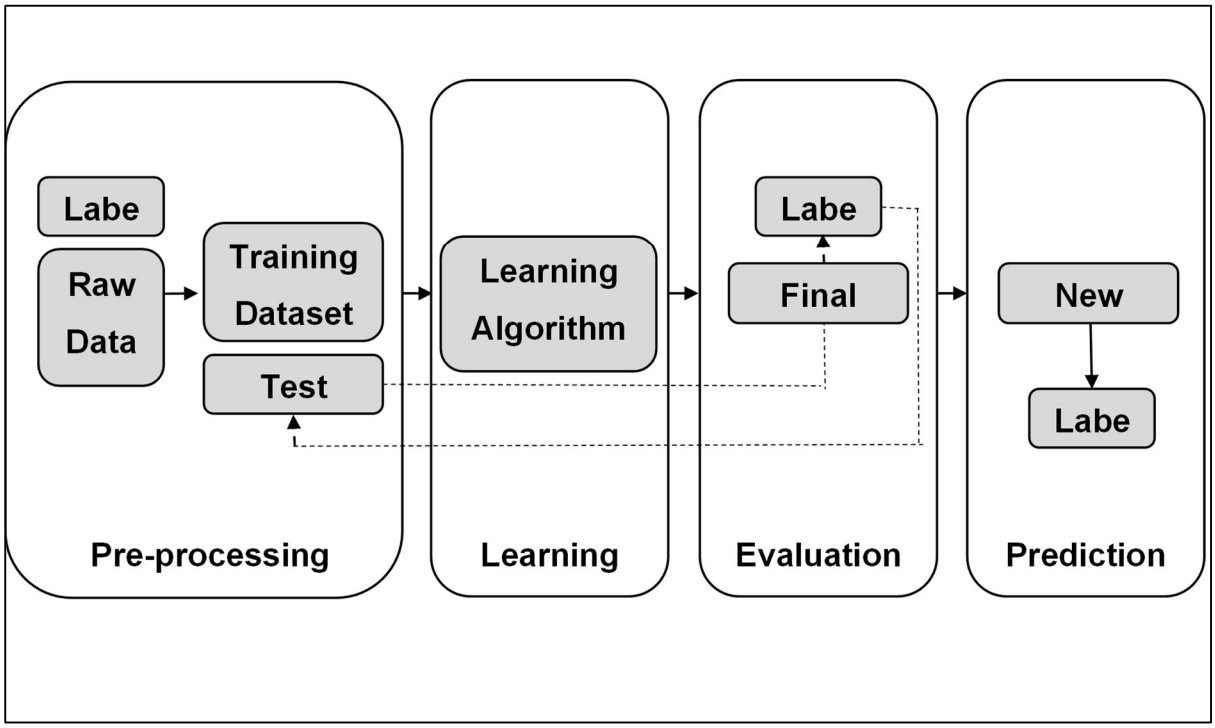

Figure 4-3 Raschka & Mirjalili's Roadmap for the Development of Machine Learning Models

Raschka & Mirjalili (2019)'s roadmap can be divided into six general steps when developing a machine learning model (Lantz, 2019; Müller & Guido, 2017). These six steps are:

1. Data Gathering,
2. Exploratory Data Analysis,
3. Data Pre-processing,
4. Model Selection,
5. Evaluation, and
6. Parameter Tuning.

4.2.1 Data Gathering

Data quality and quantity are the two most essential aspects that dictate a model's accuracy. One can ensure that these two aspects are upheld by using self-collected data. However, self-collected is not always possible. Another option would be to use pre-collected datasets, removing all the complicated aspects of self-collected data, such as strict ethics protocols. However, pre-collected datasets are not without their disadvantages. One major disadvantage is the researcher's inability to select which risk factors are collected, which may mean some factors critical to the final model are unavailable. These advantages and disadvantages were considered when deciding whether to self-collect or use a pre-collected dataset.

Data from Electronic Health Records

Datasets are typically extensive and easily analysable data sources, allowing for better predictive power. EHRs are powerful datasets used for clinical prediction models due to their size (Rajkomar, Dean, & Kohane, 2019). A benefit of EHRs is the ability to observe more metrics on more individuals at a lower cost. These benefits help predict a broader range of clinical outcomes.

Goldstein, Navar, Pencina, and Ioannidis (2017) performed an in-depth literature review to identify opportunities and challenges when developing risk prediction models with EHRs. They found that the most prevalent benefit of using EHRs is the size of the dataset. Of the 107 studies reviewed, 39 had a sample size of over 100 000 people. These large sample sizes are critical, considering that most modern risk prediction models use machine learning. Machine learning models rely on large amounts of data to produce accurate predictions. Another advantage of having large datasets is the ability to create a validation set. However, performing external validation, the use of a second dataset, was uncommon in most studies reviewed. External validation allows researchers to confidently conclude that their model will still make accurate predictions in a real-world scenario. Without external validation, ensuring the model is good at

generalising would be difficult. Generalisability is essential in machine learning, as models implemented in a real-world scenario should be able to adapt to new, previously unseen data.

Missing data is a significant problem that researchers face when using EHRs, especially in South Africa. It is widely known that South African Health Sectors struggle with poor record-keeping (Maphumulo & Bhengu, 2019). The challenge of missing data is severe, requiring careful consideration when designing the model. Secondly, it is recognised that EHRs contain, on average, more information on people with compromised health, leading to biased associations (Rusanov, Weiskopf, Wang, & Weng, 2014).

Another challenge is the potential for loss of follow-ups. When dealing with pregnancy, it is suggested that mothers have at least eight antenatal care (antenatal care) visits. Many mothers skip at least one of these, if not more (UNICEF DATA, 2021). Missing data, even for a single antenatal care visit, can cause problems with data integrity and may alter the prediction capabilities of the model. South Africa antenatal care visit statistics show that approximately 75 per cent of mothers attend four or more antenatal care visits. However, the number of mothers attending eight or more is expected to be significantly lower (Statistics South Africa, 2020). Therefore, the number of visits and intervals of visits will need to be taken into consideration. These challenges do not represent reasons not to use EHRs, but rather must be considered when deciding the type of data that the model uses.

Data from wearable sensors

With the advances in machine learning and prediction comes a need for vast amounts of data. Another way to acquire large amounts of data is by tracking live data. Currently, the most efficient way to track live data, specifically in the health sector, is to use wearable devices or wearables, as they are more commonly called. Previously, smartphones have been an easy way to collect data from patients. However, with the

mass adoption of wearables, a new opportunity has arisen that allows the sensing, collection and upload of real-time physiological data (Mobbs, Ho, Choy, Betteridge, & Lin, 2020; Seneviratne et al., 2017). This new trend has many challenges and risks that need consideration. In the article by Seneviratne et al. (2017), they consider the challenges associated with the increase in wearable devices. They specifically discuss the adoption of wearables in medical fields and mention that medical professionals still do not trust the data produced. A significant increase in the adoption of wearable devices in healthcare has been observed since the publication of Sevenviratne et al.'s (2017) paper. A reason for this was the COVID -19 pandemic. Nevertheless, any doubt that mothers and healthcare professionals have is reduced by designing the model for guidance instead of replacement.

Mixed Methods Approach

Mixing both types of data may have higher prediction possibilities. Historical data will provide data from a previous pregnancy, whereas live data will give the model real-time data on certain factors, such as BP or HR. By mixing these data types, the model can make a more informed decision since it is not limited only to the initial data collected. One consideration is that using live and historical data may require different data collection methods (De Leeuw, 2005). Combining these findings with the risk factors identified in Chapter 3, called features in this context, creates a guide for the ideal dataset for predicting preeclampsia. The ideal dataset would consist of the following features:

- DBP,
- SBP,
- Maternal age,
- BMI,
- Diabetes status,
- Hypertension History,

- Nulliparity, and
- Results of a Health Questionnaire.

The dataset would also include real-time data on features such as DBP, SBP, BMI, and the health questionnaire results.

4.2.2 Exploratory Data Analysis

As the name suggests, exploratory data analysis (EDA) is a free-form method for analysing the data. The primary aim of EDA is to draw a conclusion from the data by performing statistical and visualisation methods (Lantz, 2019; Morgenthaler, 2009). These methods include:

- Identifying the medium, mean, min and max of all numerical features to summarise the data.
- Searching for and visualising all missing and incorrect values to guide the imputation methods in the pre-processing data phase.
- Visualising and comparing the shape of the distribution for each feature.
- Generating the correlation matrix to identify existing relationships within the data.

A combination of these methods is used in Chapter 5 during each iteration's EDA and evaluation phase.

4.2.3 Data Pre-processing

Once the data has been sourced and analysed, the next phase is to prepare the data for input into the model. S. Zhang, C. Zhang and Yang (2016) describe why data preparation in machine learning is so necessary. They describe three aspects:

- Patterns may go missing due to noisy, incomplete, and inconsistent data. This is especially true for data in South Africa’s health sector (Boulle et al., 2019; Ruxwana, Herselman, & Conradie, 2010).

- Data preparation eliminates insignificant features resulting in more compact datasets. These smaller, more compact datasets still contain all the important data while allowing for quicker training times.
- Data preparation generates quality data, which leads to quality patterns.

Pre-processing, cleaning, and preparation is no easy task when dealing with complex or large datasets. Several tasks are included in the data preparation stage (Brownlee, 2020):

- Data Cleaning – mistakes or errors are identified and corrected.
- Feature Selection – the process of identifying relevant input variables.
- Data Transformation – the scale or distribution of variables is changed.
- Feature Engineering – the derivation of new variables.
- Dimension Reduction – creating compact projections of the data.

4.2.4 Model Selection

Once the data has been pre-processed and is ready to be fed into the models, the next step is to select a set of models that will be used for testing (Akinsola & J, 2017). There are two main categories of machine learning, models that use supervised or unsupervised learning. Supervised learning is defined as machine learning that uses labelled data to train algorithms. Labelled data is raw data that has been processed and tagged to identify more meaningful characteristics. An example of labelled data would be the dataset of houses with their price, suburb, colour etc. Regardless of which data collection method from Section 4.2.1 is used, all provide labelled data. Therefore, supervised learning was the selected machine learning category.

Two types of algorithms are used in supervised learning, namely regression and classification. Classification algorithms predict categorical target variables, whereas regression algorithms predict numerical target variables. The target variable is the variable that holds the outcome the model is trying to predict. The target variable for

this study will be a true or false label of whether the patient develops preeclampsia. Thus, classification algorithms were selected.

The general set of classification algorithms highlighted in the literature are the following:

- Logistic Regression
- Support Vector Machines
- K-Nearest Neighbours
- Naive Bayes
- Decision Tree
- Random Forest

4.2.5 Evaluation

After selecting and training the classifiers, an evaluation is performed to analyse the performance and shortcomings of the classifiers. Classification models are evaluated using the following metrics: Accuracy, Precision, Recall, and F1 Score (Swamynathan, 2017). Recall, also known as sensitivity, evaluates the model's ability to identify actual positives correctly (Equation 5-1). A high recall is crucial in clinical prediction models, where raising a false alarm is as important as letting the actual positive cases go undetected (Hicks et al., 2022)

$$Recall = \frac{TP}{TP + FN}$$

Equation 4-1 Recall equation for determining the model's capacity of correctly predicting a true value

The precision measures the proportion of correct positive predictions to all positive predictions (Equation 5-2).

$$P = \frac{TP}{TP + FP}$$

Equation 4-2 Precision equation for determining the quality of predictions
[63]

In binary classification, there are four possible outcomes: true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) (Figure 4-4). Figure 4-4 visualises these outcomes. TP is the number of predicted true values that were actually true, FP is the number of predicted true values that were actually false, TN is the number of predicted false values that were actually false, and FN is the number of predicted false values that were actually true.

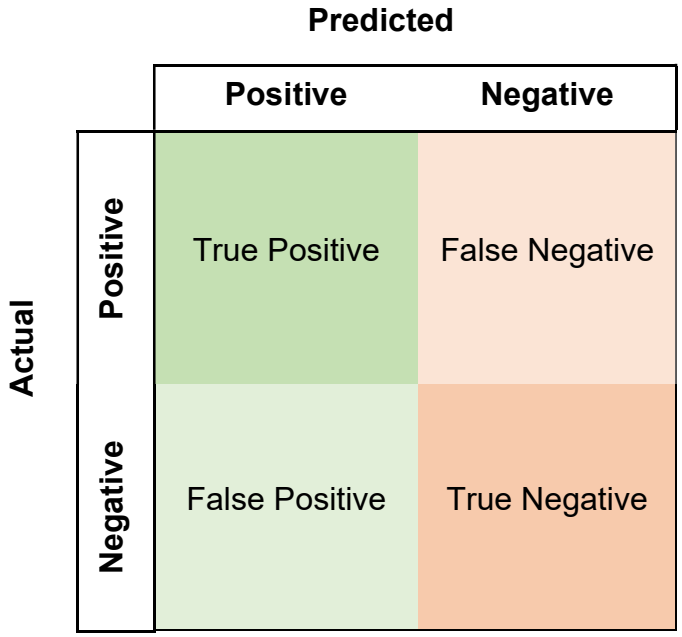

Figure 4-4 Precision-Recall Visualization (Author's work)

F-score is generally, depending on the goal of the model, a more accurate performance metric (Equation 4-3). F-score combines precision and recall to give an overall performance.

$$F = (1 + \beta)2P \cdot R\,\beta\,2 \cdot P + R$$

Equation 4-3 F-Score equation for overall performance

The last metric used is the area under the curve (AUC) for the receiver operator characteristic curve (ROC). ROC is a graph visualising the performance of classification models. AUC measures the area under the ROC curve and describes a

model's separability, the ability of the model to distinguish between positive and negative classes. An AUC of zero means the model predicates 0's as 1's and 1's as 0's, whereas an AUC of 0.5 then the model has no class separation at all.

An important method that should always be applied, specifically in studies with small datasets, is the K-Fold cross-validation (CV). K-Fold CV is used to assess the quality of classification models when the dataset does not contain enough data to include a validation set (Müller & Guido, 2017). The dataset is first shuffled and split into k equal parts. The classifier is then trained on k-1 parts and evaluated on the remaining part. A visual representation of a standard 10-Fold CV is represented in Figure 4-5. There are many different types of cross-validation. However, due to the small sample size, the Repeated Stratified KFold was used as the CV method. Using only a single run of the stratified KFold can cause a noisy estimate. Thus, the Repeated Stratified 10-Fold was used as it allows for repeating the CV, resulting in a more accurate estimate of the model's performance.

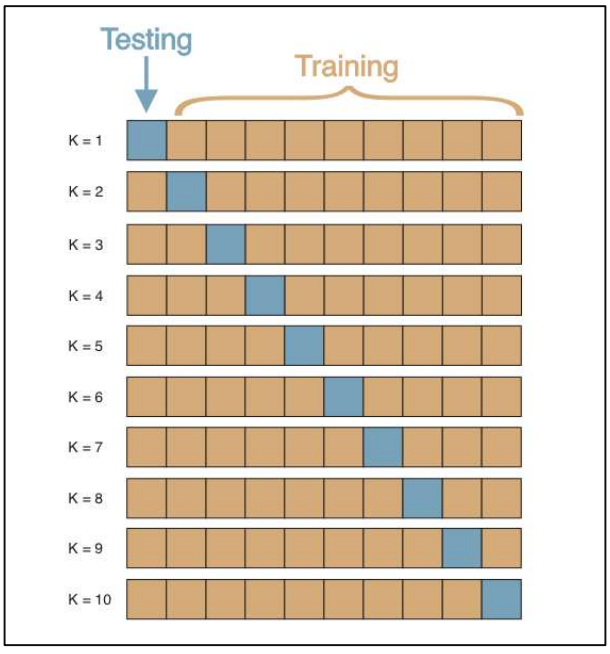

Figure 4-5 A visual representation of a 10-Fold CV **(Dantas, 2020)**

4.2.6 Hyperparameter Tuning

In machine learning, two sets of parameters are referred to when developing the model. The parameter set contains internal parameters of the model, such as weights and coefficients. In comparison, the hyperparameter set contains the model parameters defined explicitly by the developer, such as max depth or learning rate. Hyperparameter tuning is the process of searching for the optimal values of these hyperparameters. A common technique used to find the optimal values is the grid search. The grid search is a search function that, given a range of values, will test the model on every combination of parameters, returning the most optimal set (T Akinsola, Jet, & O, 2017).

4.3 Review of Existing Models

Trudell et al. (2017) developed and validated a clinical prediction model to quantify stillbirth risk. They used the Washington University School of Medicine perinatal database on singleton pregnancies from 1999-2009. The risk factors used in the model were identified through a literature review. All features with a p-value of less than 0.05 were eliminated using a backwards stepwise selection process. They developed four linear regression models predicting stillbirth at or beyond the gestational ages of 20 weeks, 24 weeks, 28 weeks, and 32 weeks. After evaluating the four models, the model that predicted stillbirths at or beyond 32 weeks was selected as the final model. Before the model was implemented in a clinical setting, Trudell et al. (2017) simplified the model by applying a point system based on the risk factors OR (Table 4.1). The referent group is depicted in Table 4.1 as Ref was assigned a score of 0. They recommended initiating antenatal testing at a score of three or more.

Although Trudell et al.'s (2017) model was relatively simple, they demonstrated another benefit of machine learning in health care. While the fields of machine learning and deep learning are making massive strides in their ability to predict complex outcomes, the prediction of stillbirths is very optimistic. However, using machine

learning to identify key risk factors that can be used in a model to assist humans in making more informed decisions is a very powerful tool.

Table 4-1 Stillbirth score card for predication of stillbirths at or beyond 32 weeks gestation, excluding foetal anomalies and aneuploidy (Trudell et al., 2017)

RISK FACTOR	OR	POINTS ASSIGNED
MATERNAL AGE (YEARS)		
≤ 18	0.42	1
19-34	Ref	0
35-39	1.24	1
≥ 40	1.55	2
BLACK	2.35	2
NULLPARITY	1.41	1
MATERNAL BMI (KG/M²)		
BMI < 25	Ref	0
BMI 25-29.9	0.98	1
BMI 30-34.9	1.75	2
BMI 35-39.9	0.55	1
BMI ≥ 40	1.17	1
SMOKING	1.25	1
CHRONIC HYPERTENSION	1.87	2
PRE-GESTATIONAL DIABETES	2.68	3
TOTAL SCORE POSSIBLE	13	

In another study, Jhee et al. (2019) developed a machine learning model to predict late-onset preeclampsia. They included 11,006 women from Yonsei University Healthcare Centre between 2005 and 2017. The data included the following features: age, height, BP, weight, gestational age, medical history, and biochemical laboratory data. The repeated-measures data such as weight, BP and laboratory data were delineated through pattern recognition and cluster analysis. Their model included a data split of 70 per cent training and 30 per cent testing, and two outcome categories, preeclampsia and no preeclampsia. Six algorithms were used for model development: Logistic Regression, Support Vector Machine, Decision Tree, Stochastic Gradient Boosting, Naïve Bayes classification, Support Vector Machine, and Random Forest. Jhee et al. (2019) used the R programming language for all the models. Pattern recognition and cluster analysis were used to evaluate each variable's influence on the predictions. The prediction model included the 14 most influential factors of the assessed variables. The Stochastic Gradient Boosting (SGB) algorithm had more than four times the highest significance. It was also the best performing model with an accuracy of 0.973 and a detection rate of 0.771.

Jhee et al. (2019) highlighted the use of mean values when dealing with prediction models in pregnancy. They agreed on the importance of taking fluctuation variability into account, incorporating repeated measured values, and including the changing patterns as an analysable factor. These recommendations support the idea of incorporating smart devices in prediction models in pregnancy. Although Jhee et al. (2019) could effectively predict late-onset preeclampsia, they faced several limitations. Many women only started the antenatal evaluation program early in the second trimester, resulting in unusable first semester data.

While concluding this study, the researcher came across a preprint of a paper published on June 9, 2022, entitled “Preeclampsia Predictor with Machine Learning: A Comprehensive and Bias-Free Machine Learning Pipeline” (Lin et al., 2022). With a similar goal, this paper was used in the evaluation of the proposed model in the final

chapters. Lin et al. (2022) used the Monitoring Mothers-to-be (nuMoM2b) dataset, which contained data from eight clinical sites across the USA (Haas et al., 2015). They selected 1758 participants, with five developing eclampsia (E), 273 who developed preeclampsia with severe features (sPE), and 1480 who had non-pregnancy induced hypertension (NPH). They identified BMI, BP, waist and neck circumference, uterine artery Doppler, diabetes, and hypertension as the most significant features. They tested several models, including Logistic Regression, Support Vector Machine, Random Forrest, and eXtreme Boosting. The dataset had four-time points: V1 (6 – 13 weeks), V2 (16 – 21), V3 (22 – 29) and V4 (delivery).

Due to the size of their dataset, they were able to split it into training, testing and validation sets with a 60-20-20 split. They used cross-fold validation with 60 folds for each model. Lin et al. compared the performance of their models for sPE+E versus NPH, and early sPE+E versus late sPE+E. Their top performing model for predicting sPE+E was the Random Forrest with an area under curve ROC of 0.63 ± 0.11 when limiting the data set to V1, 0.79 ± 0.11 when limiting the data set to V2, and 0.83 when limiting the data set to V3 (Figure 4-6). One advantage Lin et al. (2022) had was a relatively large sample size. However, compared to similar studies such as Jhee et al.'s (2017) study, which had 11,006 women, it may be considered insufficient. It may also be considered too small when looking at the number of features used. When the data was limited to V1, it had 55 features. In Chapter 3, the rule of thumb for estimating the dataset size was stated as the number of features squared. Therefore, 55 features squared would result in a dataset size of 3025 samples. This rule of thumb is used to get a rough estimate and should not be considered the optimal size. Another observation was the lack of information on any data pre-processing.

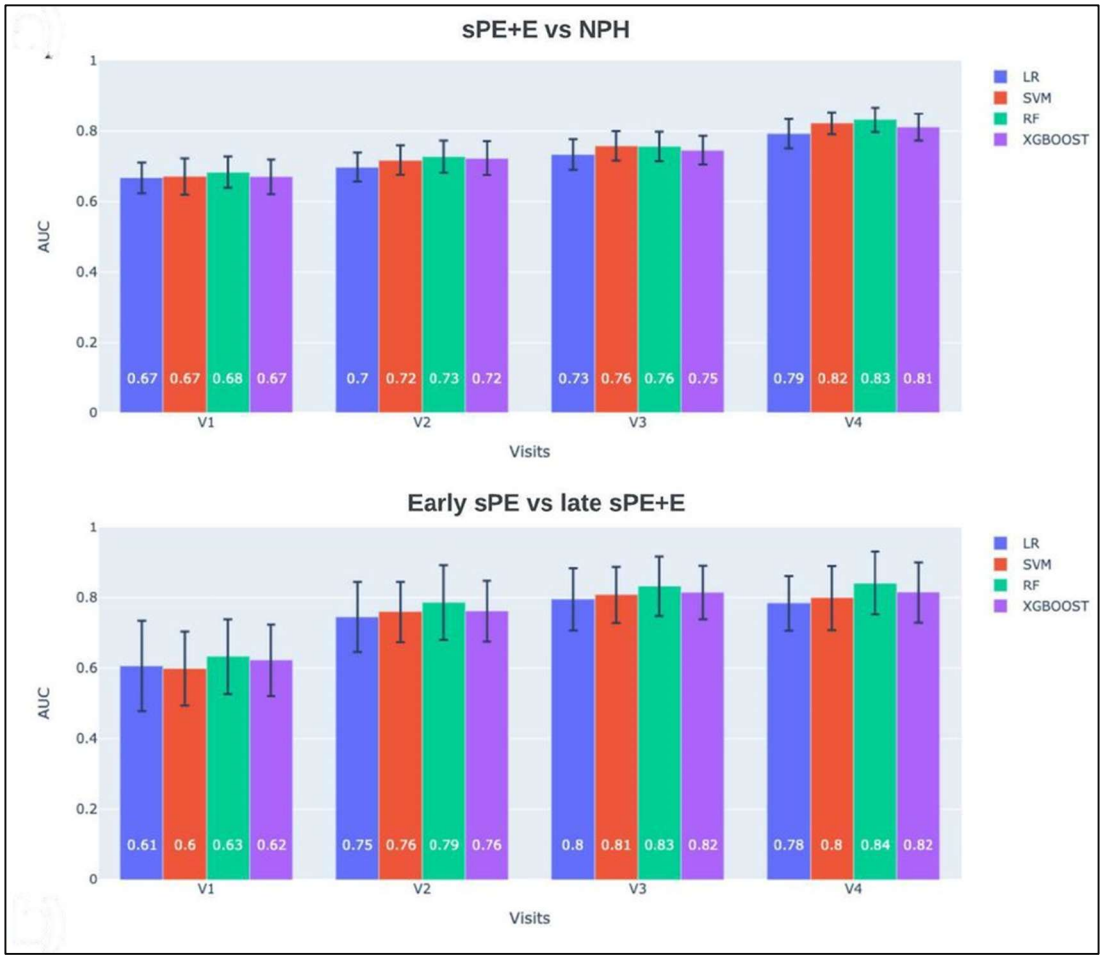

Figure 4-6 Lin et al's (2022) Model's Results

Conclusions

The aim of Chapter 4 was to answer research questions three and four:

RQ 3. What viable data source can be used to train the predictive model?

RQ 4. What existing techniques or methods can be used to predict preeclampsia?

Section 4.2 answered RQ 3 by identifying the ideal dataset for predicting preeclampsia. The ideal dataset would include the features identified in Chapter 3, with some requiring multiple readings. Section 4.3 aimed to answer RQ 4 by identifying techniques or methods for the prediction that can be used in the proposed model. The

initial set of models identified from the literature and by reviewing existing systems contained the following models:

- Logistic Regression (LR),
- Support Vector Machine (SVM),
- Naive Bayes (NB),
- K^{th} Nearest Neighbour (KNN),
- Decision Tree (DT), and
- Random Forest (RF).

These algorithms are tested in Chapter 5, where the design, development and evaluation will occur.

Chapter 5. Design and Development

5.1 Introduction

This chapter aims to answer the following research questions:

RQ 5. How can a model be designed to predict preeclampsia accurately?

RQ 6. Can the model accurately predict preeclampsia?

Section 5.2 discusses the several methods employed in the design of the proposed model. Thereafter, Sections 5.3, 5.4 and 5.5 describes the first, second and third iterations of development, respectively. Section 5.6 includes an analysis of the evaluation results for all three iterations. Finally, conclusions are provided to end the chapter in Section 5.7.

5.2 Design Process

This study combines new and existing knowledge regarding prediction in healthcare for decision-making. The following four methods were used in the design and development process for the proposed model (Figure 5-1):

- A literature review of preeclampsia – first iteration
- A literature review of prediction models in healthcare – second iteration
- An analysis of existing prediction models in healthcare, and
- Submission of a paper to the Southern African Telecommunications, Network and Communication (SATNAC) 2021 conference.

These methods fall under the ‘Awareness of problem’ and ‘Suggestion’ sections in the DSR process model. The initial designs were submitted in a full paper for the SATNAC 2021 conference, and the feedback from the reviewers served to add an iteration to discover the problem (Appendix C) further.

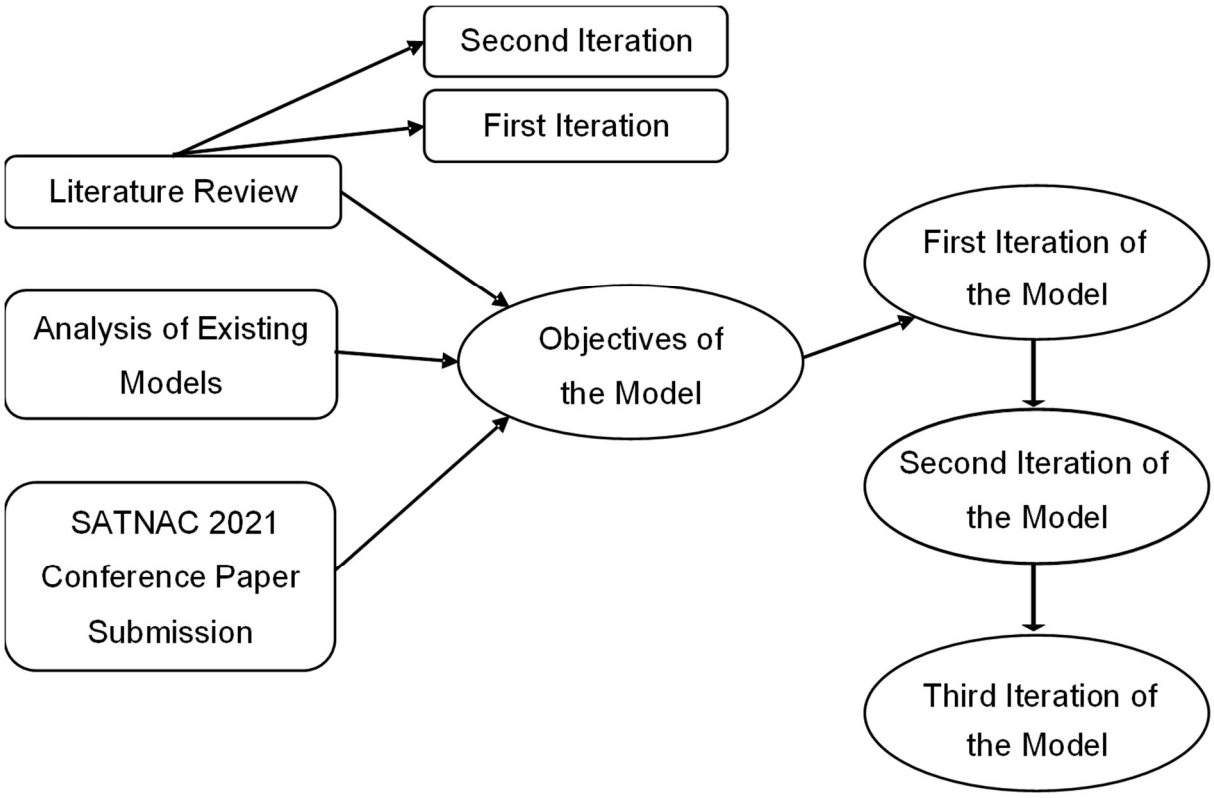

Figure 5-1 Research Methods that Contributed to the Model (Author's own)

5.3 Prediction Model for Preeclampsia (V-1)

The development of a machine learning prediction model follows the seven activities that were discussed in Section 4.2. However, an extra stage was added, namely the Setup Phase. The goal of adding a setup phase is to describe the environment and packages used in the model's development. The following stages guide the development process:

1. Setup
2. Data Gathering
3. Exploratory Data Analysis
4. Data Pre-processing
5. Model Selection

6. Evaluation
7. Hyperparameter Tuning

Iterations 2 and 3 only included activities three through seven as the data was collected in iteration 1. The models developed in these three iterations were binary classification models with the target variable being if the women had developed preeclampsia.

5.3.1 Setup Phase

Python 3.8.8 was used as the development environment with the following selected libraries, with a brief description of each:

- SciKit-Learn – is a Python library that provides machine learning algorithms such as SVMs, random forests and other supervised and unsupervised algorithms (Pedregosa et al., 2011).
- Numpy – is a fundamental Python package that provides different arrays and objects, as well as routines for faster operation on arrays (Oliphant, 2006).
- Matplotlib is a Python library that creates static, animated, and interactive visualisations (Hunter, 2007).
- Joblib – contains a method that allows the scaler used on the data to be externally stored. A scaler is a function used to change the range of the data from a feature without changing the distribution. Storing the scaler externally is important for reproducibility, as using a different scaler each time will cause inconsistent results (Joblib: Running Python Functions as Pipeline Jobs, 2008).
- Pandas – is a powerful and flexible open-source tool built on top of Python and used to analyse and manipulate data (Pandas - A Python Data Analysis Library, 2008)

5.3.2 Data Gathering

In Section 3.2, different types of data sources were examined. Initially, it was planned to perform data collection using smart devices in clinics around Port Elizabeth. However, during the initial stages of this study, the COVID-19 pandemic had just started. COVID-19 had a significant effect on the study and the planned data source. Data collection was difficult due to hospitals and clinics being closed to visitors. Thus, it was decided that an existing data source would be used. Multiple publicly available data sources were consulted to find a suitable source. However, none of the available options was suitable as they either did not have the required features or were not large enough.

Another option that would supply enough data was to contact clinics and hospitals that offer pregnancy check-ups and find out if they have databases with the patient's check-up data. The Life Hospital Group and Netcare Hospital Group were contacted and agreed to provide data after completing an application process. It was later identified that the gynaecologists and obstetricians working under these two hospital groups work independently. Thus, an application would have to be made to each gynaecologist and obstetrician. Around 30 obstetricians and gynaecologists were contacted, with only one willing to collaborate. An obstetrician with a private practice at the Mediclinic Panorama Hospital in Cape Town agreed to collaborate. He agreed to provide medical records for patients he had seen who had developed preeclampsia. He uses a large paper-based filing system for all his information storage. Thus, an online form was created, and data was entered from the patient's file into the computer. All the patients were anonymised by assigning each with a patient number. Each patient's file contained the following:

- Her pregnancy card – a pregnancy card is specific to each pregnancy for a patient and includes patient details as well as the weight and BP at each check-up
- Blood test results

- Theatre notes, and
- A variety of other documents relating to the pregnancy.

Data from two sets of 39 patients were collected who gave birth between the years of 2000 and 2022. The first set contained patients who developed preeclampsia, and the second contained those who did not. Two individuals were missing critical data and subsequently had to be removed from the study, resulting in a new sample size of 76. It must be noted that due to the limitations of the collected dataset, the results will not represent the diversity of the South African population.

5.3.3 Exploratory Data Analysis

Before processing the data, an initial data exploration (EDA) was performed. A known problem before the initial EDA was the importance 'missing' data could have on a model. Due to the low number of samples, any missing data could significantly impact the model's accuracy. Before pre-processing began, a basic understanding of the data and its shortcomings was necessary. Figure 5-2 shows the proportion of missing data per feature. These results were used during data cleaning.

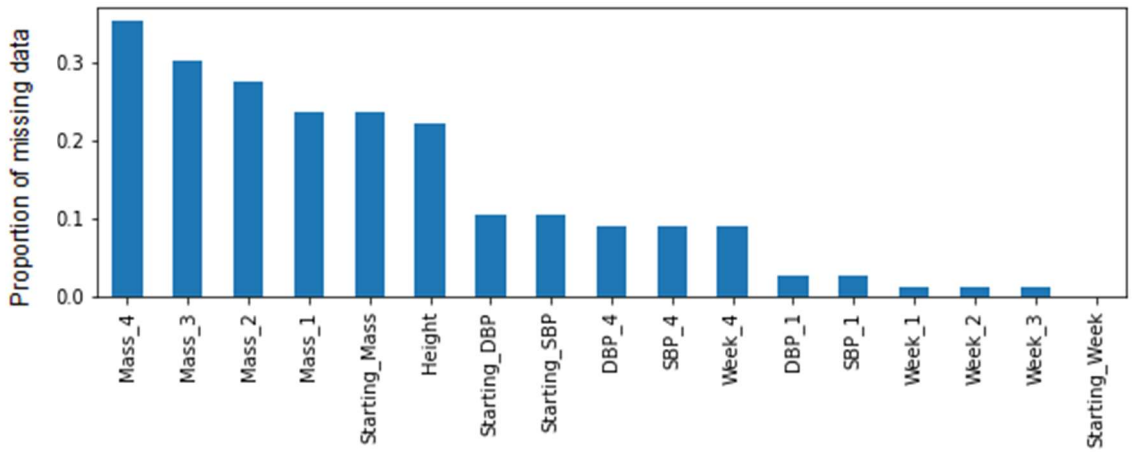

Figure 5-2 Proportions of missing data for selected features

5.3.4 Pre-processing of Data

When performing the pre-processing phase, the six tasks identified in Section 4.2.3 were performed. These tasks were:

- Data Cleaning
- Feature Selection
- Data Transformation
- Feature Engineering, and
- Dimension Reduction.

Data Cleaning

Data cleaning started with analysing the data in Excel and performing the 'eye test', where the data is searched for easily identifiable mistakes such as incorrect placement of commas, swopped fields, etc. After performing the 'eye test', an analysis of missing values was performed. It can be seen in Figure 5-2 that mass had the highest proportion of missing data.

The importance of mass in predicting preeclampsia is highly debated among experts. Fortunately, mass change during pregnancy is relatively linear (Figure 5-3). Therefore, it was easy to find a technique to estimate these masses. There are many popular data imputation techniques depending on the form and proportion of the data. Data imputation is a process in machine learning where a selected technique is used to replace missing data (Brownlee, 2020). The selected technique for replacing the missing mass values was mean imputation. Mean imputation was selected due to the linearity of mass change (Jamshidian & Mata, 2007). One problem when using mean imputation is the added bias, especially when the proportion of missing data is high. However, since five measurements for each patient were recorded, the mean could be calculated for each patient instead of the total sample (Donders, van der Heijden, Stijnen, & Moons, 2006). Only four out of the 76 patients had less than two readings;

these four patients were removed from the sample, which resulted in a new sample size of 72.

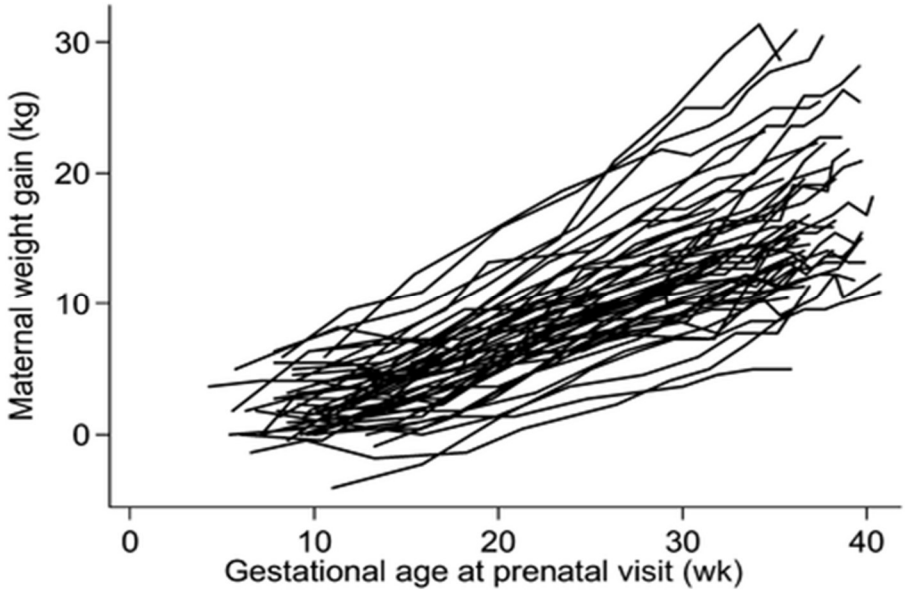

Figure 5-3 Gestational weight gain trajectories of 50 randomly selected women (Hutcheon et al., 2013)

The second highest proportion of missing values was height. With a high proportion and low sample size, there were no imputation methods that could be implemented without causing a high bias. Therefore, height was removed as a feature leaving 25 features in the feature space. The final set of features that had to be examined was the BP measurements. Unlike mass change, BP change in pregnancy is not linear; Figure 5-3 Gestational weight gain trajectories of 50 randomly selected women (Hutcheon et al., 2013) complicates data imputation. It was decided to remove the seven patients with missing BP values resulting in a sample size of 65.

The final data cleaning stage for iteration one was reviewing the categorical features and identifying rare labels. Rare labels are under-represented labels in a feature and tend to cause over-fitting. Overfitting is caused by a model that has learnt the training data too well and, as a result, loses its ability to predict the correct outcome on the test set (Hawkins, 2004). Thus, under-represented labels were encoded as ‘Rare’. Parity was the only feature identified with multiple rare labels. Figure 5-4 shows the

distribution of parity among the sample, a parity of two or more occurred significantly less than none or one. Therefore, the rare label encoder was used to encode parity of greater than two as ‘Rare’, resulting in a better distribution (Figure 5-5). The data set was now clean and could be used for feature selection.

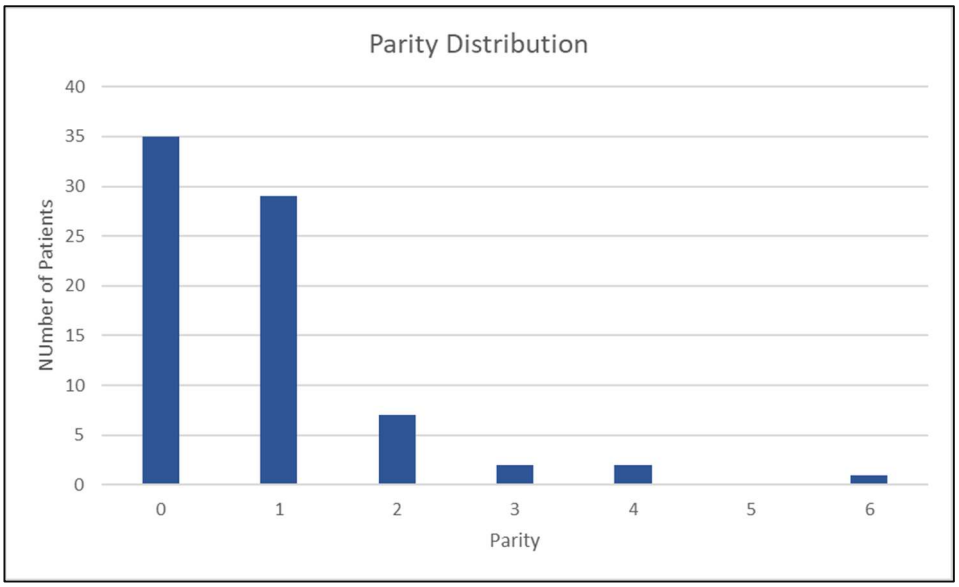

Figure 5-4 Distribution of a parity among the sample

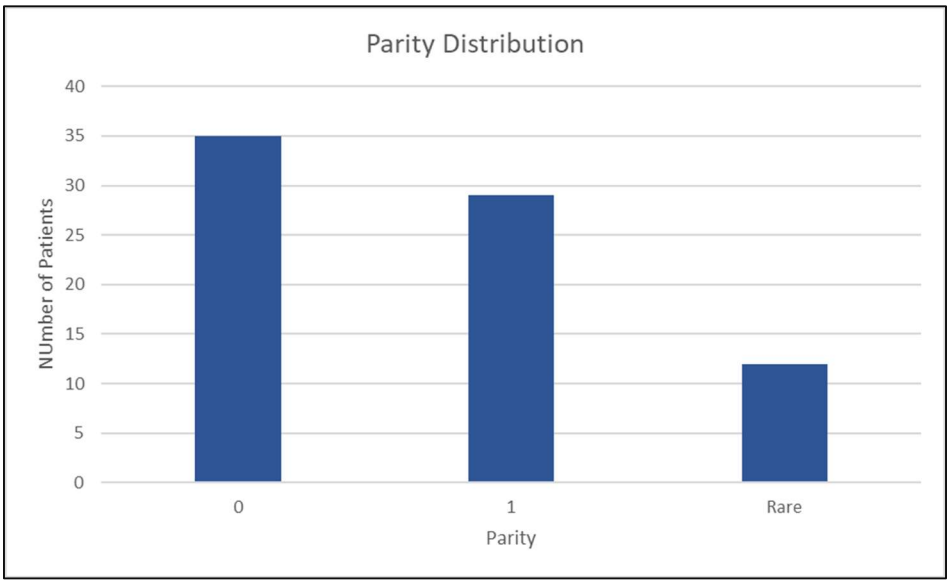

Figure 5-5 Distribution of a parity among the sample after the Rare label was encoded

Feature Selection

In Chapter 3, features of interest were selected according to the consensus of theory and practical studies about preeclampsia. Among these features were:

- Diastolic blood pressure
- Systolic blood pressure
- Maternal age
- BMI
- Diabetes status
- Hypertension history
- Nulliparity, and
- Data from a questionnaire.

Although BMI is a feature of importance, due to the number of missing height measurements, it was not included and instead replaced by mass measurements. The data from the health questionnaire was also not considered a feature due to the use of historical data where no questionnaire on patient health was completed.

Data Transformation

The only data transformation used in the first iteration was the application of a min-max scaler. Scaling data is vital as it allows the model to better understand the data in its normalised form, leading to quicker training times and higher accuracy with specific algorithms. Min-max scaling was applied to this project's model, rescaling the range of features to scale the range from -1 to 1. Scalers carry information about the data it is transforming, which could lead to overfitting. Thus, an essential rule of using a min-max scaler is to apply it separately to the train and test set.

Feature Engineering

The first iteration's feature engineering included splitting all BP readings into their separate values, namely systolic and diastolic.

Dimension Reduction

Dimensionality reduction began in iteration two as iteration one was used to get a baseline performance to compare with future iterations.

A sample of the resulting dataset after all the pre-processing was performed can be found in Appendix E.

5.3.5 Choosing a suitable model

After the data was cleaned and prepared for use by the model, the next phase was to choose a suitable model. In Chapter 4, an initial set of models was selected based on literature and a review of existing models. For reproducibility, the random state for each model was set to 0, ensuring the randomness of models was kept constant with each iteration. The initial model set comprised of the following:

1. Logistic Regression (LR)
2. Support Vector Machine (SVM)
3. Naive Bayes (NB)
4. Kth Nearest Neighbour (KNN)
5. Decision Tree (DT), and
6. Random Forest (RF).

The hyperparameters of models were all set to default for the first iteration. Models with default hyperparameters allow researchers to set a benchmark to test all future iterations. Cross-validation was also used to resample the data for the train-test sets. Cross-validation (CV) is one of the techniques used to test the effectiveness of a machine learning model; it is also a re-sampling procedure used to evaluate a model

if there is limited data. Due to the small sample size, the Repeated Stratified KFold was used as the CV method.

5.3.6 Evaluation

Table 5-2 shows the results of each evaluation method identified in Chapter 4 for each of the six models tested. LR achieved the highest accuracy, accurately predicting the outcome 79.5 per cent of the time, while also achieving the highest recall (highlighted in the red box in Table 5-2). A high recall is crucial in clinical prediction models, where raising a false alarm is as important as letting the actual positive cases go undetected (Hicks et al., 2022). The LR and SVM had good AUC values, achieving better results than most of Lin et al.'s (2022) models. AUC is used to identify model discrimination, which illustrates the iteration one's model's ability to distinguish between true and false values.

Table 5-1 Results from iteration one's models on testing data

	Accuracy	Error Rate	AUC	Precision	Recall	F1
SVM	0.776	0.224	0.776	0.834	0.718	0.748
LR	0.795	0.205	0.795	0.862	0.735	0.766
NB	0.635	0.365	0.639	0.636	0.308	0.391
KNN	0.736	0.264	0.737	0.856	0.574	0.663
DT	0.638	0.362	0.639	0.686	0.660	0.638
RF	0.747	0.253	0.747	0.815	0.708	0.726

5.3.7 Parameters Tuning

The final development phase is tuning the parameters of the selected algorithms. Sklearn's grid search was used to automatically run the algorithm through thousands of different combinations of parameter values. A grid search was done for each algorithm as most have different parameters (Appendix D). The resultant parameters set for each class selected by the grid search algorithm are listed below:

- SVM:
    - {'C': 10, 'gamma': 'scale', 'kernel': 'linear'}
- LR:
    - {'C': 0.8286427728546842, 'penalty': 'l1', 'solver': 'liblinear'}
- NB:
    - {'var_smoothing': 0.15199110829529336}
- KNN:
    - {'metric': 'manhattan', 'n_neighbors': 9}
- DT:
    - {'criterion': 'gini', 'max_depth': 7}
- RF:
    - {'criterion': 'entropy', 'max_depth': 4, 'max_features': 'sqrt'}

The resultant models were then run through the same evaluation methods used in Section 5.3.6. Although no models improved in accuracy, the LR model's recall did improve (Table 5-3). A likely reason for no noticeable increase in accuracy is due to the small size of the sample space.

Table 5-2 Comparison of iteration one's LR model before and after parameter tuning on testing dataset

	Accuracy	Error Rate	AUC	Precision	Recall	F1
$LR_{original}$	0.795	0.205	0.795	0.862	0.735	0.766
LR_{tuned}	0.785	0.215	0.785	0.831	0.753	0.765

5.4 Prediction model for preeclampsia in pregnancy (V-2)

The second iteration of the development phase started with saving the previous iteration's processed data set and using that as the starting dataset for the second iteration. Therefore, the data collection phase was skipped, and the process started with the EDA.

5.4.1 Exploratory Data Analysis

Iteration two's EDA aimed to look at the correlation matrix for all the features (Appendix F). Figure 5-6 visualises the correlation between every feature and the target variable.

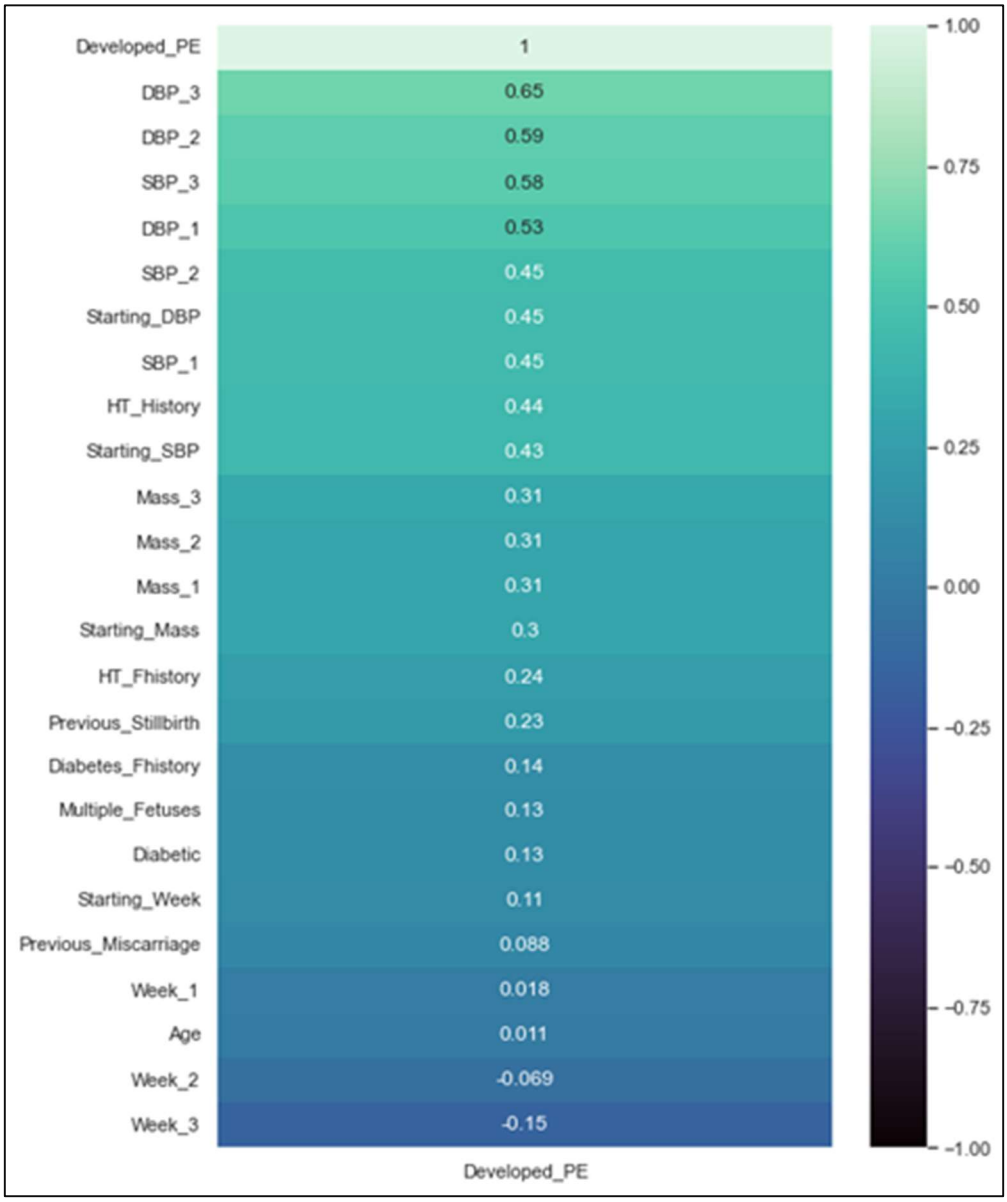

Figure 56 Features Correlating with the Development of Preeclampsia

As expected, blood pressure measurements correlate highest to the target variable. Most of the features' correlations align with the literature. One notable exception is maternal age, which has a very low correlation in this study. A reason for this could be the setting of the study, being a private obstetrician in Cape Town. A possible reason could be the obstetricians' immediate concern when a patient was older, resulting in more attention to risk factors.

5.4.2 Pre-processing of Data

As in the first iteration, the six data pre-processing activities were performed.

Data Cleaning and Feature Selection

Following on from iteration one, most of the data at this point was sufficiently cleaned. In iteration two, the data retrieved from iteration one was split into smaller subsets. The first subset consisted of only the check-up readings for the initial visit and the first check-up. Subset one is essential as it shows the potential of an early flagging prediction model. For the second subset, Random Forest was used to calculate the feature importance, acquiring the best set of features according to recursive feature elimination with cross-validation. Lastly, the third subset combined selecting the initial and first check-ups with the Random Forest feature selection.

Data Transformation

As with the first iteration, the only data transformation used in the second iteration was the application of a min-max scaler. The scaler from iteration one was saved as a .gz using the JobLib library to ensure reproducibility. JobLib is a lightweight pipeline for Python, allowing users to avoid computing the same thing more than once.

Feature Engineering

The first features to be engineered in iteration two were Previous_Miscarriage and Previous_Stillbirth. Both features had a low prevalence and correlation to developing preeclampsia (Figure 5-6). A new feature, name Previous_Mis_Still was created that combined the two features. If the patient had a previous miscarriage or a stillbirth, they would be assigned true for Previous_Mis_Still. Including this new feature allowed [76] Previous_Miscarriage and Previous_Stillbirth to be deleted, reducing the dimensionality of the datasets. Previous_Mis_Still was added to all three subsets.

Dimension Reduction

All the six subsets used lowered the dimensionality of the dataset, with the smallest having only 14 features.

The data pre-processing resulted in six new subsets to be tested:

Subset 1- Removed data for last three check-ups
Subset 2- Random forest selected features
Subset 3- Removed last three check-ups and Random Forest selected features
Subset 4- Subset 1 with the inclusion of Previous_Mis_Still
Subset 5- Subset 2 with the inclusion of Previous_Mis_Still
Subset 6- Subset 3 with the inclusion of Previous_Mis_Still

5.4.3 Choosing a suitable model

While developing the models, literature was reviewed for new models that may perform better (Müller & Guido, 2017; Schapire, 2013). Three new algorithms were identified and added to the candidate set for the second iteration to see if more complex algorithms could identify any underlying trends. The added algorithms were:

- AdaBoost
- Gradient Boosted Classifier (GBC), and
- Stochastic Gradient Descent Classifier (SGDC).

5.4.4 Evaluation

The same evaluation methods used in iteration one were used on the newly added models. The first evaluation was performed using the original dataset on the newly added models. Although there was no improvement over iteration one's LR model, the models still performed well. The SGBC was the best performer in all metrics except for precision, where there was a negligible difference compared to the GBC model. The SGBC achieved an 0.767 accuracy, 0.029 less than the LR model from iteration

1. When comparing the recall of the newly added models, it can be seen that all achieved recalls are comparable to the top three models from iteration 1 (Table 5-4). These models were added to the set of models for future iterations.

Table 5-3 Recall values for iteration one's and two's models on the testing dataset

	SVM	LR	NB	KNN	DT	RF	SGBC	GBC	Ada
Recall	0.724	0.753	0.628	0.558	0.66	0.708	0.73	0.693	0.712

The next phase was evaluating all the models on each of the six subsets generated from the pre-processing. Although most of the results were worse, one interesting finding was the SVM model on subset 3 (highlighted in the red box in Table 5-5). This subset only contained the first two check-ups and the newly added feature. The mean gestation week for the first check-up was the 16th week. This is significant as it means the model can pick up, within the first 16 weeks, if a woman is likely to develop preeclampsia 70 per cent of the time.

Table 5-4 Results of all models on the six subsets on the testing dataset

Subset 1

	Acc	Err	AUC	Prec	Rec	F1
SVM	0.708	0.292	0.708	0.783	0.56	0.62
LR	0.665	0.335	0.666	0.742	0.481	0.553
NB	0.708	0.292	0.708	0.836	0.511	0.592
KNN	0.681	0.319	0.68	0.775	0.44	0.531
DT	0.66	0.34	0.658	0.706	0.642	0.647
RF	0.688	0.312	0.689	0.729	0.688	0.675
SGBC	0.692	0.308	0.692	0.709	0.676	0.65
GBC	0.698	0.302	0.699	0.758	0.681	0.68
Ada	0.561	0.439	0.56	0.572	0.611	0.566

Subset 2

	Acc	Err	AUC	Prec	Rec	F1
SVM	0.755	0.245	0.753	0.79	0.711	0.715
LR	0.735	0.265	0.735	0.771	0.708	0.715
NB	0.754	0.246	0.756	0.878	0.601	0.683
KNN	0.721	0.279	0.722	0.888	0.492	0.601
DT	0.665	0.335	0.665	0.72	0.667	0.654
RF	0.763	0.237	0.763	0.847	0.708	0.739
SGBC	0.719	0.281	0.719	0.732	0.763	0.713
GBC	0.752	0.248	0.752	0.804	0.719	0.734
Ada	0.683	0.317	0.682	0.73	0.679	0.671

Subset 3

	Acc	Err	AUC	Prec	Rec	F1
SVM	0.724	0.276	0.724	0.753	0.701	0.707
LR	0.665	0.335	0.667	0.679	0.668	0.641
NB	0.718	0.282	0.719	0.822	0.544	0.616
KNN	0.651	0.349	0.651	0.751	0.414	0.503
DT	0.66	0.34	0.658	0.72	0.642	0.648
RF	0.693	0.307	0.693	0.723	0.686	0.673
SGBC	0.689	0.311	0.69	0.688	0.678	0.65
GBC	0.704	0.296	0.705	0.747	0.701	0.69
Ada	0.579	0.421	0.579	0.606	0.578	0.568

Subset 4

	Acc	Err	AUC	Prec	Rec	F1
SVM	0.713	0.287	0.713	0.781	0.558	0.619
LR	0.665	0.335	0.666	0.742	0.481	0.553
NB	0.702	0.298	0.702	0.783	0.588	0.631
KNN	0.681	0.319	0.68	0.775	0.44	0.531
DT	0.628	0.372	0.626	0.671	0.636	0.622
RF	0.687	0.313	0.688	0.716	0.66	0.651
SGBC	0.69	0.31	0.693	0.696	0.651	0.633
GBC	0.652	0.348	0.652	0.679	0.647	0.633
Ada	0.564	0.436	0.563	0.576	0.606	0.563

Subset 5

	Acc	Err	AUC	Prec	Rec	F1
SVM	0.767	0.233	0.766	0.798	0.726	0.727
LR	0.735	0.265	0.735	0.771	0.708	0.715
NB	0.767	0.233	0.768	0.881	0.643	0.712
KNN	0.724	0.276	0.725	0.888	0.497	0.606
DT	0.663	0.337	0.664	0.709	0.667	0.655
RF	0.752	0.248	0.751	0.826	0.708	0.732
SGBC	0.719	0.281	0.718	0.753	0.744	0.712
GBC	0.773	0.227	0.772	0.836	0.733	0.752
Ada	0.68	0.32	0.679	0.721	0.685	0.671

Subset 6

	Acc	Err	AUC	Prec	Rec	F1
SVM	0.681	0.319	0.681	0.709	0.638	0.65
LR	0.665	0.335	0.667	0.679	0.668	0.641
NB	0.721	0.279	0.722	0.809	0.604	0.653
KNN	0.662	0.338	0.663	0.769	0.436	0.523
DT	0.657	0.343	0.655	0.7	0.636	0.642
RF	0.669	0.331	0.67	0.704	0.66	0.648
SGBC	0.707	0.293	0.706	0.722	0.7	0.675
GBC	0.696	0.304	0.697	0.73	0.697	0.682
Ada	0.583	0.417	0.583	0.602	0.596	0.577

5.4.5 Parameter Tuning

The final phase in the development of the model was to tune the parameters of the selected algorithms. The resultant parameters set for each model selected by the grid search algorithm are listed below:

- SGBC:
  - {'loss': 'perceptron', 'penalty': 'l2'}
- GBC:
  - {'learning_rate': 0.02, 'max_depth': 6, 'n_estimators': 20, 'subsample': 0.2}
- AdaBoost:
  - {'algorithm': 'SAMME', 'learning_rate': 0.01, 'n_estimators': 30}

The resultant models were then run through the evaluation methods. The results for the three new models are shown in Table 5-6. Although no significant increases were observed over most of the metrics, there was a significant increase in Recall for the SGDC and GBC algorithms. When comparing these recall values to recall values in Table 5-5, it can be seen that the SGDC model is now the best performing model in terms of recall.

Table 5-5 Results for newly added model's parameter tuned on the original testing dataset from iteration one

	Accuracy	Error Rate	AUC	Precision	Recall	F1
SGDC	0.765	0.235	0.764	0.793	0.771	0.746
GBC	0.749	0.251	0.749	0.797	0.735	0.738
Ada	0.725	0.275	0.724	0.841	0.611	0.676

5.5 Prediction model for preeclampsia (V-3)

The final iteration of the development phase used the first iteration's dataset and iteration two's subsets, with the last three check-ups removed as the initial dataset for iteration 3.

5.5.1 Exploratory Data Analysis

Categorical representation was analysed and represented by using bar graphs (Figure 5-8). Multiple_Fetuses, Previous_Miscarriage, Previous_Stillbirth, Diabetic and Diabetes_FHistory had their true labels heavily under-represented. Under-represented labels can cause a significant loss in performance, specifically in small datasets (Afrose et al., 2021). Unfortunately, the only way to deal with under-represented labels is to over-sample, which was not an option in this study, and data simulation, which was out of the scope of this study. Therefore, the only option was to discard them.

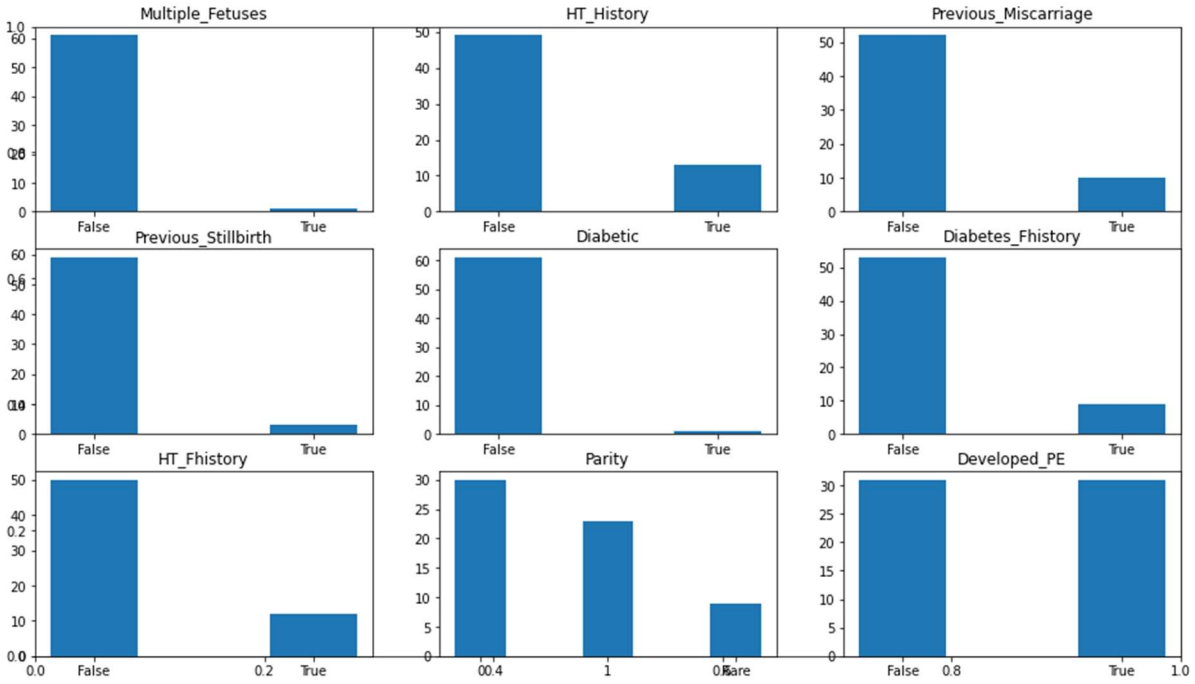

Figure 5-6 Categorical distribution visual representation (n=72)

5.5.2 Pre-processing of Data

As in the first iteration, the six data pre-processing activities were performed.

Data Cleaning and Dimension Reduction

The only data cleaning performed in this iteration was deleting the five features identified in the EDA. These were:

- Multiple_Fetuses
- Previous_Miscarriage
- Previous_Stillbirth
- Diabetic, and
- Diabetes_FHistory.

Together with these changes was the use of subset one, where the dimensionality was already reduced. Therefore, the two subsets for iteration three were the following:

- Subset 7 - Original dataset with the five features identified in Section 5.5.1 removed
- Subset 8 - Subset one with the five features identified in Section 5.5.1 removed.

Data Transformation

The min-max scaler used in iterations one and two were used to ensure reproducibility.

Feature Selection

No new feature selection method was used in this iteration.

Feature Engineering

No feature engineering was performed.

5.5.3 Choosing a suitable model

A new *ensemble* model was added to the existing pool of models for iteration three. An ensemble model combines multiple classification models to increase the model's accuracy as well as its generalisability. The ensemble model used three of the previously tested models, including SVM, LR, and AdaBoost. A voting method was used as the researcher was familiar with the method. The weighting of each model's votes was also adjusted.

5.5.4 Evaluation

All three models were tested on subsets seven and eight, using the same evaluation methods as iterations one and two. The following results were obtained for this iteration's subset one (Table 5-7). Noticeable increases in accuracy were observed across most models. The LR model continues to be the best performing model in terms of accuracy. However, a notable change was the SVM model, which achieved the highest Recall, albeit only by 0.01 compared to the parameter-tuned LR model from iteration one. All models' AUC also increased, showing that removing the insignificant features helped the models become less discriminatory.

Table 5-6 Results for all models on subset seven testing dataset

	Accuracy	Error Rate	AUC	Precision	Recall	F1
SVM	0.828	0.172	0.828	0.872	0.783	0.828
LR	0.838	0.162	0.838	0.908	0.772	0.838
NB	0.777	0.223	0.775	0.827	0.739	0.777
KNN	0.794	0.206	0.795	0.868	0.712	0.794
DT	0.677	0.323	0.677	0.715	0.665	0.677
RF	0.752	0.248	0.751	0.803	0.724	0.752
SGBC	0.737	0.263	0.734	0.748	0.722	0.737
GBC	0.707	0.293	0.706	0.744	0.707	0.707
Ada	0.719	0.281	0.717	0.744	0.743	0.719
Combined	0.788	0.212	0.790	0.863	0.701	0.788

Subset seven achieved an increase in accuracy among all iteration two's models tuned on the subsets where the last three check-ups were discarded (Table 5-8). However, a significant decrease in Recall and an increase in Precision were observed. In other words, the models made better quality predictions, but were not good at identifying if a woman would develop preeclampsia (Hicks et al., 2022). This is something to consider for future research when implementing the proposed model in a clinical setting.

Table 5-7 Results for all models on iteration three's subset two testing dataset

	Accuracy	Error Rate	AUC	Precision	Recall	F1
SVM	0.733	0.267	0.733	0.802	0.604	0.658
LR	0.773	0.227	0.774	0.879	0.643	0.712
NB	0.619	0.381	0.622	0.586	0.275	0.352
KNN	0.676	0.324	0.677	0.743	0.574	0.615
DT	0.685	0.315	0.684	0.731	0.668	0.675
RF	0.698	0.302	0.699	0.747	0.681	0.678
SGBC	0.664	0.336	0.665	0.675	0.661	0.627
GBC	0.669	0.331	0.669	0.719	0.649	0.657
Ada	0.586	0.414	0.586	0.599	0.619	0.584
Combined	0.717	0.283	0.717	0.809	0.562	0.631

5.5.5 Parameter Tuning

The final phase in the development of the model was to tune the hyperparameters of the selected algorithms. A grid search was done for each algorithm as, generally, the hyperparameters are different for each one (Appendix D). The resultant hyperparameters set for each model selected by the grid search algorithm are listed below:

- SVM:
  - {'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
- LR:
  - {'C': 3.727593720314938, 'penalty': 'l2', 'solver': 'newton-cg'}
- NB:
  - {'var_smoothing': 0.1}

- KNN:
    - {'metric': 'manhattan', 'n_neighbors': 5}
- DT:
    - {'criterion': ‘entropy’, 'max_depth': 50}
- RF:
    - {'criterion': 'entropy', 'max_depth': 3, 'max_features': 'log2'}
- SGBC:
    - {'loss': 'log', 'penalty': 'l1'}
- GBC:
    - {'learning_rate': 0.03, 'max_depth': 8, 'n_estimators': 20, 'subsample': 0.5}
- AdaBoost:
    - {'algorithm': 'SAMME.R', 'learning_rate': 0.1, 'n_estimators': 2}

The resultant models were then run through the same evaluation methods. A few models had minor increases in accuracy but nothing significant. Subset seven’s highest performing model was the AdaBoost, which saw an increase of almost 20 per cent in accuracy and 15 per cent in Recall, making it the best performing model for the datasets limited to the first check-up (Table 5-9).

Table 5-8 Results for all parameter tuned models on iteration three's subset two testing dataset

	Accuracy	Error Rate	AUC	Precision	Recall	F1
SVM	0.746	0.254	0.746	0.844	0.576	0.652
LR	0.757	0.243	0.756	0.861	0.631	0.692
NB	0.759	0.241	0.760	0.907	0.551	0.655
KNN	0.687	0.313	0.688	0.754	0.544	0.605
DT	0.685	0.315	0.684	0.731	0.668	0.675
RF	0.700	0.300	0.700	0.732	0.690	0.677
SGBC	0.640	0.360	0.641	0.648	0.635	0.603
GBC	0.673	0.327	0.672	0.731	0.646	0.653
Ada	0.770	0.230	0.770	0.811	0.765	0.762
Combined	0.724	0.276	0.724	0.844	0.533	0.622

5.6 Discussion of Findings and the Final Model

After the evaluation of the models in Chapter 5, several discussion points arose, which are discussed in this section. The first discussion point relates to the data collection and the dataset used. Due to the COVID-19 pandemic, which will be further discussed under Section 6.5, acquiring data was extremely difficult for all researchers. The dataset collected was much smaller than expected. This created a knock-on effect that could be seen in the development phase, where options on data cleaning, feature engineering etc., were limited. The difficulty in sourcing large amounts of data is a considerable challenge for researchers in machine learning and deep learning. Open data policies in research are being pushed in many countries, but applying these policies will still take many years (Wu, Moylan, Inman, & Graf, 2019).

A second discussion point is comparing the significance of risk factors from this study to the risk factors identified in the literature review. In Jhee et al.'s (2017) and Li et al.'s (2022) studies, SBP had the highest importance by a significant margin compared to this study, where DBP had higher importance than SBP. Another interesting finding was the very low significance of maternal age in this study. As mentioned in Section 5.4.1, this is possibly due to the private practice chosen as the data source. Diabetes also had an unusually low significance, although this is due to the underrepresentation of true values. The increase in accuracy from iteration two to three supports this claim as it was removed in iteration 3.

Iteration one's results were promising from the beginning, with the best performing model being the logistic regression (LR) model achieving an accuracy of 0.795 and a recall of 0.735. The LR model had all default hyperparameters and used the original dataset. Iteration two added three new machine learning algorithms to the model pool. The only positive change resulting from iteration two was the SGDC model, which achieved the highest recall of 0.771. Iteration two also added six new subsets as data sources with no noticeable increase in performance. One positive insight from iteration two was the results from subset 3, where the subset was limited to the initial and first check-ups. The SVM model achieved an accuracy of 0.742, while having a good recall of 0.701. This result showed the proposed model's potential to predict preeclampsia much earlier in the pregnancy than expected.

The final iteration, iteration three, is where the final and best performing models were observed. All models were tested on two new datasets, where all insignificant features were eliminated and a focus on minimising dimensionality was prioritised. The SVM model achieved a notable 0.835 accuracy and a recall of 0.767 on subset seven. This was the best performing model of the whole study and is one of two models considered as the artefact of this study. The second model was the AdaBoost on subset eight, which achieved an accuracy of 0.77 and a recall of 0.765. Considering the proposed model has only been exposed to half of the data the SVM had on subset seven, the

scores are very similar. As mentioned in Section 5.4.4, the mean gestation week for the first check-up is 16.1. If a model like this is implemented in a rural clinic situation, a model that can identify high-risk patients within the first 16 weeks is very beneficial. Implementing such a model will allow resources to be distributed to only those in need in resource-scarce settings.

Comparing these results to the similar existing models identified in Section 4.3 will indicate how these models measure up to other models. Jhee et al.'s (2017) model was trained on a dataset approximately 150 times the size of this study's dataset. Their study was looking at predicting late-onset preeclampsia, where they achieved an overall accuracy of 0.973. A more equivalent comparison would be this study's models to Lin et al.'s (2022) top-performing model, the RF model. In this study, the SVM model on subset seven was compared to Lin et al.'s (2022) RF model; the accuracy achieved was equal to 0.835 and 0.83, respectively. However, comparing this study's AdaBoost model on subset eight was an interesting finding with Lin et al.'s (2022) V1 model. Both models can predict, within the first 16 weeks of pregnancy, if a woman is at risk of developing preeclampsia. This study's Adaboost model achieved an accuracy of 0.77 compared to Lin et al.'s (2022) study of 0.63 for their V1 model. Their V1 model is from week 6 to 13. A better comparison would be this study's model, with a model between V1 and V2 (0.68 – 0.79). These results are comparable even though this study only had 68 samples compared to Lin et al.'s 3000 samples.

The potential of the models suggested in this dissertation can be seen when considering the setting where the model would be implemented, where any help would be better than none.

5.7 Conclusions

This chapter aimed to answer the following RQs:

RQ 5. How should a model be designed to accurately predict preeclampsia?
RQ 6. Can the model accurately predict preeclampsia?

Chapter 4 identified a development cycle for the development of a machine learning model. This development cycle followed the six phases of development identified in Chapter 4. These phases were:

1. Data Gathering
2. Exploratory Data Analysis
3. Data Pre-processing
4. Model Selection
5. Evaluation, and
6. Hyperparameter Tuning.

Chapter 5 aimed to implement the identified development cycle and develop a model to predict the development of preeclampsia. Following this development cycle resulted in the two final models, the SVM model trained on data available before the 16th week and the AdaBoost trained on data from the first three check-ups. Both final models achieved similar or better results than those identified in the literature. Therefore, validating the development cycle used and answering RQ 5. These results confirmed the model's ability to accurately predict preeclampsia accurately, answering RQ 6.

Chapter 6. Conclusions

6.1 Introduction

This chapter will discuss the findings, make recommendations for future research, and conclude the study with some closing remarks. The main research objective of this study was:

Develop a machine learning model that can support doctors and midwives in identifying patients at risk of developing preeclampsia.

Two models were developed that could accurately predict the development of preeclampsia, one before 16 weeks' gestation and the other within three check-ups. These models were developed using knowledge acquired from the existing literature and an analysis of existing models. The research objectives and questions are evaluated in Section 6.3. Section 6.4 highlights the contributions of the study, followed by the limitations of the study in Section 6.5. The chapter ends with several recommendations for future research in Section 6.6 and a summary of the research in Section 6.7.

6.2 Achievement of Research Objectives and Questions

The effectiveness of this research can be evaluated if the research objectives were achieved. A real-world problem that affected many women in South Africa during a period that should be filled with joy was identified. This research aimed to solve that problem by developing a prediction model that can assist overworked health professionals by identifying high-risk patients early in pregnancy. Table 6-1 summarises if and how each research objective was achieved.

Table 6-1 Review of Research Objectives and Achievements

Research Objective		Description	Source
RO1	Investigate if preeclampsia is a significant risk for pregnant women.	Several risks for pregnant women were identified, but preeclampsia was identified as the most prevalent condition.	Section 3.3
RO2	Identify the factors that may affect a woman's chances of developing preeclampsia during pregnancy.	Eight factors were identified that affect the development of preeclampsia. These eight factors were: DBP, SBP, Maternal age, BMI, Diabetes status, Hypertension History, Nulliparity, Health Questionnaire data	Section 3.4 and 3.5
RO3	Identify a viable data source to train the predictive model.	Two viable datasets were identified depending on the progress of Covid-19. The two types were health records or a field study with smart devices.	Section 4.2
RO4	Identify existing techniques or methods that can be used to predict preeclampsia.	Several techniques were discussed, but a machine learning model was selected as the preferred model to predict whether a woman is at risk of developing preeclampsia.	Section 4.2 and 4.3
RO5	Investigate if a model can be designed to accurately predict preeclampsia during pregnancy.	In Chapter 5, multiple models were developed. The SVM and the Adaboost models were selected as the final models to predict if a woman is at risk of developing preeclampsia.	Chapter 5

RO6	Evaluate if the model is successful in accurately predicting preeclampsia.	When comparing the model to similar models, it was clear that the proposed model was successful in predicting preeclampsia.	Section 5.6

6.3 Contributions of the Study

The contribution of this research can be split into two categories: theoretical and practical contributions. The theoretical contribution revolves around machine learning research as a valuable tool in resource-limited areas. The practical contribution is a model that can be applied in a clinical environment.

6.3.1 Theoretical Contributions

This study had the following theoretical contributions:

- Show the potential of machine learning models in healthcare (Section 4.2 and 4.2)
- An accredited conference publication in the Proceedings of SATNAC 2021.

A thorough literature review and analysis of existing literature on machine learning models in healthcare, specifically pregnancy monitoring, was performed. Several findings were identified and formed the guidelines for developing machine learning models. These guidelines were then validated in Chapter 5, where they were used to guide the researcher on the development of the model. The conference publication extended this by providing a proof-of-concept for developing a model to predict preeclampsia.

To the researcher's knowledge, this proposed model is the first of its kind in South Africa and the first to use a South African antenatal data set.

6.3.2 Practical Contributions

The first practical contribution is the identification of eight critical factors for predicting preeclampsia through a literature review. These factors were:

- DBP,
- SBP,
- Maternal age,
- BMI,
- Diabetes status,
- Hypertension History,
- Nulliparity, and
- Results of a Questionnaire.

Unfortunately, due to the data capture, some of these identified factors were unable to be used. BMI and the use of a questionnaire were not used in the final model. These risk factors are complex enough to allow for accurate predictions while not requiring complex measuring devices.

The other practical contribution of this study is the final two models, namely the SVM and AdaBoost. These two models can be implemented in different live settings with minor modifications. The main goal of the model is not to replace the obstetricians or midwives but to affirm their decisions, specifically when the health facilities are overworked and understaffed.

6.4 Problems and Limitations

Starting this research in 2020, one month before the COVID-19 pandemic was declared, made several aspects of this study more difficult. Travel restrictions and social distancing halted this study for six months until they were lifted. Following the lifting of the strict travelling restrictions, all hospitals were closed to most research and

clinical trials if the research did not focus on COVID-19. This did not even consider the emotional and social effects of the pandemic on everyone.

These effects resulted in the first limitation, namely the size of the dataset. Due to the reasons listed above and electronic medical records not being readily available in South Africa, the only dataset that could be sourced was a small paper-based dataset from private practice in Cape Town. As mentioned throughout this study, the larger the dataset, the better the results would have been.

6.5 Recommendations for Future Research

The first recommendation for future research is aimed at the digitisation and accessibility of medical records and data for researchers in South Africa. Lin et al. (2022) used a dataset containing multiple clinics' data with hundreds of recorded vitals and characteristics. Madhi et al.'s (2019) study identified many factors affecting the women of South Africa and highlighted the need for researchers to look for solutions to these problems. However, research requires data; and unfortunately, South Africa is far behind where we should be regarding digitisation.

The accessibility of medical data has been and continues to be the most significant setback when it comes to machine learning in healthcare (Rajkomar, Dean, & Kohane, 2019) and is why it needs to be a priority for future research. The second recommendation is related to COVID-19 and the long-term effects we observed after more than two years of it being active. The long-term effects of COVID-19 may completely change the risk factors associated with preeclampsia and therefore need to be researched.

6.6 Summary

DSRM was successfully applied in this study to allow for the research questions to be answered and an artefact, the proposed model, to be produced. The findings revealed a promising model that can predict if a pregnant woman is at risk of developing preeclampsia within the first 16 weeks of pregnancy. However, the findings highlighted the downfalls of using a small sample set. Three iterations were performed to develop the two chosen models presented in this chapter as the final artefact. To the researcher's knowledge, these models are first machine learning models developed for predicting preeclampsia using a South African dataset. However, for these models to be implemented properly and effectively, the systems we use to capture patients' data in the private and public sectors will need to be upgraded and staff trained. This will be resource-intensive and financially expensive but will promote better research into the areas affecting South African women. No woman should ever experience the tragic loss of stillbirth if it could be predicted beforehand.

References

Abalos, E., Cuesta, C., Grosso, A. L., Chou, D., & Say, L. (2013). Global and regional estimates of preeclampsia and eclampsia: A systematic review. In *European Journal of Obstetrics and Gynecology and Reproductive Biology* (Vol. 170, Issue 1, pp. 1–7). Elsevier Ireland Ltd. https://doi.org/10.1016/j.ejogrb.2013.05.005

Afrose, S., Song, W., Nemeroff, C. B., Lu, C., Yao, D., & Author, C. (2021). Overcoming Underrepresentation in Clinical Datasets for Accurate Subpopulation-specific Prognosis. *MedRxiv*, 2021.03.26.21254401. https://doi.org/10.1101/2021.03.26.21254401

Airoldi, J., & Weinstein, L. (2007). Clinical significance of proteinuria in pregnancy. *Obstetrical and Gynecological Survey*, *62*(2), 117–124. https://doi.org/10.1097/01.OGX.0000253301.55009.AC

Akbari, S., Khodadadi, B., Ahmadi, S. A. Y., Abbaszadeh, S., & Shahsavar, F. (2018). Association of vitamin D level and vitamin D deficiency with risk of preeclampsia: A systematic review and updated meta-analysis. *Taiwanese Journal of Obstetrics & Gynecology*, *57*(2), 241–247. https://doi.org/10.1016/J.TJOG.2018.02.013

Alonzo, T. A. (2009). Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating: By Ewout W. Steyerberg. *American Journal of Epidemiology*, *170*(4), 2. https://doi.org/10.1093/aje/kwp129

Aminu, M., Unkels, R., Mdegela, M., Utz, B., Adaji, S., & van den Broek, N. (2014). Causes of and factors associated with stillbirth in low- and middle-income countries: a systematic literature review. *BJOG : An International Journal of Obstetrics and Gynaecology*, *121*, 141–153. https://doi.org/10.1111/1471-0528.12995

An, C., Lim, H., Kim, D. W., Chang, J. H., Choi, Y. J., & Kim, S. W. (2020). Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. *Scientific Reports 2020 10:1*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-75767-2

Ananth, C. V., Keyes, K. M., & Wapner, R. J. (2013a). Pre-eclampsia rates in the United States, 1980-2010: age-period-cohort analysis. *BMJ (Clinical Research Ed.)*, *347*. https://doi.org/10.1136/BMJ.F6564

Ananth, C. V., Keyes, K. M., & Wapner, R. J. (2013b). Pre-eclampsia rates in the United States, 1980-2010: Age-period-cohort analysis. *BMJ (Online)*, *347*. https://doi.org/10.1136/bmj.f6564

Askie, L. M., Duley, L., Henderson-Smart, D. J., & Stewart, L. A. (2007). Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data. *Lancet*, *369*(9575), 1791–1798. https://doi.org/10.1016/S0140-6736(07)60712-0

Autopilot AI | Tesla. (n.d.). Retrieved June 17, 2021, from https://www.tesla.com/autopilotAI

Ayo-Yusuf, L. (2015). *South Africa needs a new way to address the doctor shortage*. https://theconversation.com/south-africa-needs-a-new-way-to-address-the-doctor-shortage-41136

Bartsch, E., Medcalf, K. E., Park, A. L., Ray, J. G., Al-Rubaie, Z. T. A., Askie, L. M., Berger, H., Blake, J., Graves, L., Kingdom, J. C., Lebovic, G., Lord, S. J., Maguire, J. L., Mamdani, M. M., Meloche, J., Urquia, M. L., & Van Wagner, V. (2016). Clinical risk factors for pre-eclampsia determined in early pregnancy: Systematic review and meta-analysis of large cohort studies. *The BMJ*, *353*. https://doi.org/10.1136/bmj.i1753

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design Science Research Contributions: Finding a Balance between Artifact and Theory. *Journal of the Association for Information Systems*, *19*(5), 358–376. https://doi.org/10.17705/1jais.00495

Battarbee, A. N., Sinkey, R. G., Harper, L. M., Oparil, S., & Tita, A. T. N. (2020). Chronic Hypertension in Pregnancy. *American Journal of Obstetrics and Gynecology*. https://doi.org/10.1016/j.ajog.2019.11.1243

Berwick, D. M., Nolan, T. W., & Whittington, J. (2017). The Triple Aim: Care, Health, And Cost. *Https://Doi.Org/10.1377/Hlthaff.27.3.759*, *27*(3), 759–769. https://doi.org/10.1377/HLTHAFF.27.3.759

Bezerra, P. C. F. M., Leão, M. D., Queiroz, J. W., Melo, E. M. D., Pereira, F. V. M., Nóbrega, M. H., Jeronimo, A. K., Ferreira, L. C., Jerônimo, S. M. B., & De Araújo, A. C. P. F. (2010). Family history of hypertension as an important risk factor for the development of severe preeclampsia. *Acta Obstetricia et Gynecologica Scandinavica*, *89*(5), 612–617. https://doi.org/10.3109/00016341003623720

Bhardwaj, R., Nambiar, A. R., & Dutta, D. (2017). A Study of Machine Learning in Healthcare. *Proceedings - International Computer Software and Applications Conference*, *2*, 236–241. https://doi.org/10.1109/COMPSAC.2017.164

Births and Deaths Registration Act 51 of 1992, (1992). https://www.gov.za/sites/default/files/gcis_document/201409/a511992.pdf

Bottaci, L., Drew, P. J., Hartley, J. E., Hadfield, M. B., Farouk, R., Lee, P. W. R., MacIntyre, I. M. C., Duthie, G. S., & Monson, J. R. T. (1997). Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *Lancet*, *350*(9076), 469–472. https://doi.org/10.1016/S0140-6736(96)11196-X

Boulle, A., Heekes, A., Tiffin, N., Smith, M., Mutemaringa, T., Zinyakatira, N., Phelanyane, F., Pienaar, C., Buddiga, K., Coetzee, E., van Rooyen, R., Dyers,

R., Fredericks, N., Loff, A., Shand, L., Moodley, M., de Vega, I., & Vallabhjee, K. (2019). Data Centre Profile: The Provincial Health Data Centre of the Western Cape Province, South Africa. *International Journal of Population Data Science*, *4*(2). https://doi.org/10.23889/IJPDS.V4I2.1143

Brantsæter, A. L., Haugen, M., Samuelsen, S. O., Torjusen, H., Trogstad, L., Alexander, J., Magnus, P., & Meltzer, H. M. (2009). A Dietary Pattern Characterized by High Intake of Vegetables, Fruits, and Vegetable Oils Is Associated with Reduced Risk of Preeclampsia in Nulliparous Pregnant Norwegian Women. *The Journal of Nutrition*, *139*(6), 1162. https://doi.org/10.3945/JN.109.104968

Bresnick, J. (2017). *Google's Machine Learning, Imaging Analytics Flag Breast Cancer*. https://healthitanalytics.com/news/googles-machine-learning-imaging-analytics-flag-breast-cancer

Brown, M. A., Magee, L. A., Kenny, L. C., Karumanchi, S. A., McCarthy, F. P., Saito, S., Hall, D. R., Warren, C. E., Adoyi, G., & Ishaku, S. (2018). Hypertensive disorders of pregnancy: ISSHP classification, diagnosis, and management recommendations for international practice. In *Hypertension* (Vol. 72, Issue 1, pp. 24–43). Lippincott Williams and Wilkins. https://doi.org/10.1161/HYPERTENSIONAHA.117.10803

Brown, P. (2018). *Noncommunicable diseases: Risk factors*. https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/ncd-risk-factors

Brownlee, J. (2020). *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python* (v1.2). Machine Learning Mastery, 2020. https://books.google.co.za/books?hl=en&lr=&id=uAPuDwAAQBAJ&oi=fnd&pg=PP1&dq=data+preparation+techniques+machine+learning&ots=Cl3FsbaQxR&sig=Vi-juWPicuxR2FrZhFgANgs8Cs0#v=onepage&q=data preparation techniques machine learning&f=false

Bujold, E., Roberge, S., Lacasse, Y., Bureau, M., Audibert, F., Marcoux, S., Forest, J. C., & Giguère, Y. (2010). Prevention of preeclampsia and intrauterine growth restriction with aspirin started in early pregnancy: A meta-analysis. In *Obstetrics and Gynecology* (Vol. 116, Issue 2, pp. 402–414). https://doi.org/10.1097/AOG.0b013e3181e9322a

Burgen, M. (2018). *Google AI Could Help Predict When You Will Die*. Bloomberg. https://www.bloomberg.com/news/articles/2018-06-18/google-is-training-machines-to-predict-when-a-patient-will-die#xj4y7vzkg

Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo, J. L., Jones, D. W., Materson, B. J., Oparil, S., Wright, J. T., & Roccella, E. J. (2003).

The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: The JNC 7 Report. In *Journal of the American Medical Association* (Vol. 289, Issue 19, pp. 2560–2572). American Medical Association. https://doi.org/10.1001/jama.289.19.2560

Clarke, M. (2022). *Critical shortage of doctors in SA*. https://www.da.org.za/2022/05/critical-shortage-of-doctors-in-sa-less-than-1-doctor-for-every-1-000-patients

Cunningham, P., Cord, M., & Delany, S. J. (2008). Supervised learning. *Cognitive Technologies*, 21–49. https://doi.org/10.1007/978-3-540-75171-7_2

Dantas, J. (2020). *The importance of k-fold cross-validation for model prediction in machine learning | by Jaime Dantas | Towards Data Science*. Towards Data Science. https://towardsdatascience.com/the-importance-of-k-fold-cross-validation-for-model-prediction-in-machine-learning-4709d3fed2ef

De Bernis, L., Kinney, M. V., Stones, W., Ten Hoope-Bender, P., Vivio, D., Leisher, S. H., Bhutta, Z. A., Gülmezoglu, M., Mathai, M., Belizán, J. M., Franco, L., McDougall, L., Zeitlin, J., Malata, A., Dickson, K. E., & Lawn, J. E. (2016). Stillbirths: Ending preventable deaths by 2030. *The Lancet*, *387*(10019), 703–716. https://doi.org/10.1016/S0140-6736(15)00954-X

De Leeuw, E. D. (2005). *To Mix or Not to Mix Data Collection Modes in Surveys*. www.websm.org

Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising qualitative and quantitative evidence: A review of possible methods. *Journal of Health Services Research and Policy*, *10*(1), 45–53. https://doi.org/10.1258/1355819052801804

Donders, A. R. T., van der Heijden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, *59*(10), 1087–1091. https://doi.org/10.1016/J.JCLINEPI.2006.01.014

Ebonwu, J., Mumbauer, A., Uys, M., Wainberg, M. L., & Medina-Marino, A. (2018). Determinants of late antenatal care presentation in rural and peri-urban communities in South Africa: A cross-sectional study. *PLoS ONE*, *13*(3). https://doi.org/10.1371/journal.pone.0191903

Finlayson, K., & Downe, S. (2013). Why Do Women Not Use Antenatal Services in Low- and Middle-Income Countries? A Meta-Synthesis of Qualitative Studies. *PLoS Medicine*, *10*(1). https://doi.org/10.1371/JOURNAL.PMED.1001373

Frizzell, J. D., Liang, L., Schulte, P. J., Yancy, C. W., Heidenreich, P. A., Hernandez, A. F., Bhatt, D. L., Fonarow, G. C., & Laskey, W. K. (2017). Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: Comparison of

machine learning and other statistical approaches. *JAMA Cardiology*, *2*(2), 204–209. https://doi.org/10.1001/jamacardio.2016.3956

Gardosi, J., Madurasinghe, V., Williams, M., Malik, A., & Francis, A. (2013). Maternal and fetal risk factors for stillbirth: Population based study. *BMJ (Online)*, *346*(7893). https://doi.org/10.1136/bmj.f108

Garovic, V. D., & Hayman, S. R. (2007). Hypertension in pregnancy: an emerging risk factor for cardiovascular disease. *Nature Clinical Practice. Nephrology*, *3*(11), 613–622. https://doi.org/10.1038/ncpneph0623

Glaros, C., & I. Fotiadis, D. (2005). *Wearable Devices in Healthcare* (pp. 237–264). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11311966_8

Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. A. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *Journal of the American Medical Informatics Association*, *24*(1), 198–208. https://doi.org/10.1093/jamia/ocw042

Haas, D. M., Parker, C. B., Wing, D. A., Parry, S., Grobman, W. A., Mercer, B. M., Simhan, H. N., Hoffman, M. K., Silver, R. M., Wadhwa, P., Iams, J. D., Koch, M. A., Caritis, S. N., Wapner, R. J., Esplin, M. S., Elovitz, M. A., Foroud, T., Peaceman, A. M., Saade, G. R., … Reddy, U. M. (2015). A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). *American Journal of Obstetrics and Gynecology*, *212*(4), 539.e1-539.e24. https://doi.org/10.1016/J.AJOG.2015.01.019

Haddad, B., Deis, S., Goffinet, F., Paniel, B. J., Cabrol, D., & Sibaï, B. M. (2004). Maternal and perinatal outcomes during expectant management of 239 severe preeclamptic women between 24 and 33 weeks' gestation. *American Journal of Obstetrics and Gynecology*, *190*(6), 1590–1595. https://doi.org/10.1016/J.AJOG.2004.03.050

Harper, L., Kalfa, N., Beckers, G. M. A., Kaefer, M., Nieuwhof-Leppink, A. J., Fossum, M., Herbst, K. W., & Bagli, D. (2020). The impact of COVID-19 on research. *Journal of Pediatric Urology*, *16*(5), 715–716. https://doi.org/10.1016/J.JPUROL.2020.07.002

Hauger, M. S., Gibbons, L., Vik, T., & Belizán, J. M. (2008). Prepregnancy weight status and the risk of adverse pregnancy outcome. *Acta Obstetricia et Gynecologica Scandinavica*, *87*(9), 953–959. https://doi.org/10.1080/00016340802303349

Hawkins, D. M. (2004). The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences*, *44*(1), 1–12. https://doi.org/10.1021/CI0342472

Hevner, A., & Chatterjee, S. (2010). Design Science Research in Information Systems. *Design Research in Information Systems*, *22*, 9–22. https://doi.org/10.1007/978-

1-4419-5653-8

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly: Management Information Systems*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports 2022 12:1*, *12*(1), 1–9. https://doi.org/10.1038/s41598-022-09954-8

Hinton, G. (2018). Deep learning-a technology with the potential to transform health care. In *JAMA - Journal of the American Medical Association* (Vol. 320, Issue 11, pp. 1101–1102). American Medical Association. https://doi.org/10.1001/jama.2018.11100

Hofmeyr, G. J., & Mentrop, L. (2015). Time for 'basic antenatal care plus' in South Africa?'. In *South African Medical Journal* (Vol. 105, Issue 11, pp. 902–903). South African Medical Association. https://doi.org/10.7196/SAMJ.2015.v105i11.10186

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(03), 90–95. https://doi.org/10.1109/MCSE.2007.55

Hutcheon, J. A., Platt, R. W., Abrams, B., Himes, K. P., Simhan, H. N., & Bodnar, L. M. (2013). A weight-gain-for-gestational-age z score chart for the assessment of maternal weight gain in pregnancy. *The American Journal of Clinical Nutrition*, *97*(5), 1062–1067. https://doi.org/10.3945/AJCN.112.051706

Hypertension in pregnancy. (2013). *Obstetrics and Gynecology*, *122*(5), 1122–1131. https://doi.org/10.1097/01.AOG.0000437382.03963.88

"Iterative Learning-InsideAIML." (n.d.). Retrieved June 15, 2021, from https://insideaiml.com/blog/Iterative-Learning-292

Jamshidian, M., & Mata, M. (2007). Advances in Analysis of Mean and Covariance Structure when Data are Incomplete. *Handbook of Latent Variable and Related Models*, 21–44. https://doi.org/10.1016/B978-044452044-9/50005-7

Jhee, J. H., Lee, S., Park, Y., Lee, S. E., Kim, Y. A., Kang, S. W., Kwon, J. Y., & Park, J. T. (2019). Prediction model development of late-onset preeclampsia using machine learning-based methods. *PLoS ONE*, *14*(8), e0221202. https://doi.org/10.1371/journal.pone.0221202

Joblib: running Python functions as pipeline jobs. (2008). https://joblib.readthedocs.io/en/latest/

Johannesson, P., & Perjons, E. (2014). An introduction to design science. *An Introduction to Design Science*, *9783319106328*, 1–197.

https://doi.org/10.1007/978-3-319-10632-8

Jones, D., & Gregor, S. (2007). The Anatomy of a Design Theory. *Journal of the Association for Information Systems*, *8*(5), 19. https://doi.org/10.17705/1jais.00129

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. In *Science* (Vol. 349, Issue 6245, pp. 255–260). American Association for the Advancement of Science. https://doi.org/10.1126/science.aaa8415

Khan, K. S., Wojdyla, D., Say, L., Gülmezoglu, A. M., & Van Look, P. F. (2006). WHO analysis of causes of maternal death: a systematic review. *Lancet (London, England)*, *367*(9516), 1066–1074. https://doi.org/10.1016/S0140-6736(06)68397-9

Lamminpää, R., Vehviläinen-Julkunen, K., Gissler, M., & Heinonen, S. (2012). Preeclampsia complicated by advanced maternal age: a registry-based study on primiparous women in Finland 1997-2008. *BMC Pregnancy and Childbirth*, *12*(1), 1–5. https://doi.org/10.1186/1471-2393-12-47/TABLES/2

Lantz, B. (2013). *Machine Learning with R* (1st ed., Vol. 1). Packt Publishing.

Lantz, B. (2019). *Machine Learning with R: Expert techniques for predictive modeling* (Third). Packt Publishing. https://books.google.co.za/books?hl=en&lr=&id=iNuSDwAAQBAJ&oi=fnd&pg=PP1&dq=how+to+perform+machine+learning+models&ots=O86Xxd5CQ-&sig=DwHnMIFUdPprPXomzpqez4uy0mA#v=onepage&q=how to perform machine learning models&f=false

Lawn, J. E., Blencowe, H., Waiswa, P., Amouzou, A., Mathers, C., Hogan, D., Flenady, V., Frøen, J. F., Qureshi, Z. U., Calderwood, C., Shiekh, S., Jassir, F. B., You, D., McClure, E. M., Mathai, M., Cousens, S., Kinney, M. V., De Bernis, L., Heazell, A., … Draper, E. S. (2016). Stillbirths: Rates, risk factors, and acceleration towards 2030. *The Lancet*, *387*(10018), 587–603. https://doi.org/10.1016/S0140-6736(15)00837-5

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In *Nature* (Vol. 521, Issue 7553, pp. 436–444). Nature Publishing Group. https://doi.org/10.1038/nature14539

Lindheimer, M. D., Taler, S. J., & Cunningham, F. G. (2008). Hypertension in pregnancy. *Journal of the American Society of Hypertension*, *2*(6), 484–494. https://doi.org/10.1016/J.JASH.2008.10.001

Madazli, R., Yuksel, M. A., Imamoglu, M., Tuten, A., Oncul, M., Aydin, B., & Demirayak, G. (2014). Comparison of clinical and perinatal outcomes in early- and late-onset preeclampsia. *Archives of Gynecology and Obstetrics*, *290*(1), 53–57. https://doi.org/10.1007/S00404-014-3176-X

Madhi, S. A., Briner, C., Maswime, S., Mose, S., Mlandu, P., Chawana, R., Wadula, J., Adam, Y., Izu, A., & Cutland, C. L. (2019). Causes of stillbirths among women from South Africa: a prospective, observational study. *The Lancet Global Health*, *7*(4), e503–e512. https://doi.org/10.1016/S2214-109X(18)30541-2

Maphumulo, W. T., & Bhengu, B. R. (2019). Challenges of quality improvement in the healthcare of South Africa post-apartheid: A critical review. In *Curationis* (Vol. 42, Issue 1). NLM (Medline). https://doi.org/10.4102/curationis.v42i1.1901

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2

Maternal mortality. (n.d.). Retrieved June 15, 2021, from https://www.who.int/news-room/fact-sheets/detail/maternal-mortality

Mobbs, R. J., Ho, D., Choy, W. J., Betteridge, C., & Lin, H. (2020). COVID-19 is shifting the adoption of wearable monitoring and telemedicine (WearTel) in the delivery of healthcare: opinion piece. *Annals of Translational Medicine*, *8*(20), 1285–1285. https://doi.org/10.21037/atm-20-3678

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning, second edition - Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar - Google Books*. The MIT Press. https://books.google.co.za/books?hl=en&lr=&id=dWB9DwAAQBAJ&oi=fnd&pg=PR5&dq=machine+learning&ots=AyoQTSr2r0&sig=uA9coMnaohutimIV6uq0tSFo7ww#v=onepage&q&f=false

Morgenthaler, S. (2009). Exploratory data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(1), 33–44. https://doi.org/10.1002/WICS.2

Müller, A. C., & Guido, S. (2017). *Introduction to Machine Learning with Python: A Guide for Data Scientists* (First). O'Reilly Media. https://books.google.co.za/books?hl=en&lr=&id=1-4IDQAAQBAJ&oi=fnd&pg=PP1&dq=machine+learning+guide&ots=28IUKKJMTW&sig=P3FZdGLp2hTGqEY-QxoNktJ3Q-g#v=onepage&q=machine learning guide&f=false

Myer, L., & Harrison, A. (2003). Why do womn seek antenatal care late? Perspectives from rural South Africa. *Journal of Midwifery and Women's Health*, *48*(4), 268–272. https://doi.org/10.1016/S1526-9523(02)00421-X

Myers, J. E., Myatt, L., Roberts, J. M., & Redman, C. W. G. (2019). COLLECT, a collaborative database for pregnancy and placental research studies worldwide. In *BJOG: An International Journal of Obstetrics and Gynaecology* (Vol. 126, Issue 1, pp. 8–10). Blackwell Publishing Ltd. https://doi.org/10.1111/1471-0528.15393

Myers, M. D., & Venable, J. R. (2014). A set of ethical principles for design science

research in information systems. *Information & Management*. https://doi.org/10.1016/j.im.2014.01.002

Nathan, H. L., Seed, P. T., Hezelgrave, N. L., De Greeff, A., Lawley, E., Conti-Ramsden, F., Anthony, J., Steyn, W., Hall, D. R., Chappell, L. C., & Shennan, A. H. (2018). Maternal and perinatal adverse outcomes in women with pre-eclampsia cared for at facility-level in South Africa: A prospective cohort study. *Journal of Global Health*, *8*(2). https://doi.org/10.7189/jogh.08.020401

NICE. (2019). *Hypertension in pregnancy: diagnosis and management*. www.nice.org.uk/guidance/ng133

North, R. A., McCowan, L. M. E., Dekker, G. A., Poston, L., Chan, E. H. Y., Stewart, A. W., Black, M. A., Taylor, R. S., Walker, J. J., Baker, P. N., & Kenny, L. C. (2011). Clinical risk prediction for pre-eclampsia in nulliparous women: development of model in international prospective cohort. *BMJ*, *342*(7803). https://doi.org/10.1136/BMJ.D1875

Oliphant, T. E. (2006). *Guide to NumPy*. http://www.trelgol.com

Pandas - A Python Data Analysis Library. (2008). https://pandas.pydata.org/about/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhfer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., & Duchesnay, M. P. E. (2011). Scikit-learn: Machine Learning in Python. In *Journal of Machine Learning Research* (Vol. 12). http://scikit-learn.sourceforge.net.

Peffers, K., Tuunanen, T., & Gengler, C. (2006). The design science research process: a model for producing and presenting information systems research. *Journal of Management Information Systems*, *February*, 83–106.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pries-Heje, J., Baskerville, R., & Venable, J. R. (2008). *Strategies for Design Science Research Evaluation*. *87*. http://aisel.aisnet.org/ecis2008/87

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, *380*(14), 1347–1358. https://doi.org/10.1056/NEJMRA1814259

Reinebrant, H., Leisher, S., Coory, M., Henry, S., Wojcieszek, A., Gardener, G., Lourie, R., Ellwood, D., Teoh, Z., Allanson, E., Blencowe, H., Draper, E., Erwich, J., Frøen, J., Gardosi, J., Gold, K., Gordijn, S., Gordon, A., Heazell, A., … Flenady, V. (2018). Making stillbirths visible: a systematic review of globally

reported causes of stillbirth. *BJOG: An International Journal of Obstetrics & Gynaecology*, *125*(2), 212–224. https://doi.org/10.1111/1471-0528.14971

Rezk, M., Gamal, A., & Emara, M. (2015). Maternal and fetal outcome in de novo preeclampsia in comparison to superimposed preeclampsia: a two-year observational study. *Hypertension in Pregnancy*, *34*(2), 137–144. https://doi.org/10.3109/10641955.2014.982329

Roberts, J. M., & Gammill, H. S. (2005). Preeclampsia: Recent insights. In *Hypertension* (Vol. 46, Issue 6, pp. 1243–1249). Lippincott Williams & Wilkins. https://doi.org/10.1161/01.HYP.0000188408.49896.c5

Rosser, M. L., & Katz, N. T. (2013). Preeclampsia: An Obstetrician's Perspective. In *Advances in Chronic Kidney Disease* (Vol. 20, Issue 3, pp. 287–296). W.B. Saunders. https://doi.org/10.1053/j.ackd.2013.02.005

Rusanov, A., Weiskopf, N. G., Wang, S., & Weng, C. (2014). Hidden in plain sight: Bias towards sick patients when sampling patients with sufficient electronic health record data for research. *BMC Medical Informatics and Decision Making*, *14*(1), 1–9. https://doi.org/10.1186/1472-6947-14-51

Ruxwana, N. L., Herselman, M. E., & Conradie, D. P. (2010). ICT applications as e-health solutions in rural healthcare in the Eastern Cape Province of South Africa. *The HIM Journal*, *39*(1), 17–26. https://doi.org/10.1177/183335831003900104

Savaj, S., & Vaziri, N. (2012). *An overview of recent advances in pathogenesis and diagnosis of preeclampsia.* https://escholarship.org/uc/item/3863p1k7

Seneviratne, S., Hu, Y., Nguyen, T., Lan, G., Khalifa, S., Thilakarathna, K., Hassan, M., & Seneviratne, A. (2017). A Survey of Wearable Devices and Challenges. In *IEEE Communications Surveys and Tutorials* (Vol. 19, Issue 4, pp. 2573–2620). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/COMST.2017.2731979

Shen, M., Smith, G. N., Rodger, M., White, R. R., Walker, M. C., & Wen, S. W. (2017). Comparison of risk factors and outcomes of gestational hypertension and pre-eclampsia. *PLoS ONE*, *12*(4). https://doi.org/10.1371/JOURNAL.PONE.0175914

Statistics South Africa. (2020). *Maternal Health Indicators*. http://www.statssa.gov.za/?p=13100

Swamynathan, M. (2017). Mastering Machine Learning with Python in Six Steps. In *Mastering Machine Learning with Python in Six Steps*. Apress. https://doi.org/10.1007/978-1-4842-2866-1

Sylvia Patience Ngxongo, T. (2019). Basic Antenatal Care Approach to Antenatal Care Service Provision. In *Selected Topics in Midwifery Care*. IntechOpen. https://doi.org/10.5772/intechopen.79361

T Akinsola, J. E., Jet, A., & O, H. J. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology*, *48*. https://doi.org/10.14445/22312803/IJCTT-V48P126

Tranquilli, A. L., Dekker, G., Magee, L., Roberts, J., Sibai, B. M., Steyn, W., Zeeman, G. G., & Brown, M. A. (2014). The classification, diagnosis and management of the hypertensive disorders of pregnancy: A revised statement from the ISSHP. In *Pregnancy Hypertension* (Vol. 4, Issue 2, pp. 97–104). Elsevier B.V. https://doi.org/10.1016/j.preghy.2014.02.001

Trochim, W. M. K. (2006). *Research Design - Research Methods Knowledge Base*. https://conjointly.com/kb/research-design/

Trudell, A. S., Tuuli, M. G., Colditz, G. A., Macones, G. A., & Odibo, A. O. (2017). A stillbirth calculator: Development & internal validation of a clinical prediction model to quantify stillbirth risk. *PLoS ONE*, *12*(3), 1–13. https://doi.org/10.1371/journal.pone.0173461

Tucker, K. L., Bankhead, C., Hodgkinson, J., Roberts, N., Stevens, R., Heneghan, C., Rey, É., Lo, C., Chandiramani, M., Taylor, R. S., North, R. A., Khalil, A., Marko, K., Waugh, J., Brown, M., Crawford, C., Taylor, K. S., Mackillop, L., & McManus, R. J. (2018). How do home and clinic blood pressure readings compare in pregnancy? A systematic review and individual patient data meta-analysis. In *Hypertension* (Vol. 72, Issue 3, pp. 686–694). Lippincott Williams and Wilkins. https://doi.org/10.1161/HYPERTENSIONAHA.118.10917

Tucker, K. L., Taylor, K. S., Crawford, C., Hodgkinson, J. A., Bankhead, C., Carver, T., Ewers, E., Glogowska, M., Greenfield, S. M., Ingram, L., Hinton, L., Khan, K. S., Locock, L., Mackillop, L., McCourt, C., Pirie, A. M., Stevens, R., & McManus, R. J. (2017). Blood pressure self-monitoring in pregnancy: Examining feasibility in a prospective cohort study. *BMC Pregnancy and Childbirth*, *17*(1), 442. https://doi.org/10.1186/s12884-017-1605-0

UNICEF DATA. (2021). *Antenatal care*. https://data.unicef.org/topic/maternal-health/antenatal-care/

Vaishnavi, V., & Kuechler, B. (2004). *Design Research in Information Systems*. http://www.desrist.org/design-research-in-information-systems/

van de Leur, R. R., Boonstra, M. J., Bagheri, A., Roudijk, R. W., Sammani, A., Taha, K., Doevendans, P. A. F. M., van der Harst, P., van Dam, P. M., Hassink, R. J., van Es, R., & Asselbergs, F. W. (2020). Big data and artificial intelligence: Opportunities and threats in electrophysiology. *Arrhythmia and Electrophysiology Review*, *9*(3), 146–154. https://doi.org/10.15420/AER.2020.26

van den Heuvel, J. F. M., Kariman, S. S., van Solinge, W. W., Franx, A., Lely, A. T., & Bekker, M. N. (2019). SAFE@HOME – Feasibility study of a telemonitoring

platform combining blood pressure and preeclampsia symptoms in pregnancy care. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, *240*, 226–231. https://doi.org/10.1016/j.ejogrb.2019.07.012

van der Merwe, A., Gerber, A., & Smuts, H. (2017). Mapping a design science research cycle to the postgraduate research report. *Communications in Computer and Information Science*, *730*, 293–308. https://doi.org/10.1007/978-3-319-69670-6_21

Venable, J., Pries-Heje, J., & Baskerville, R. (2012). A Comprehensive Framework for Evaluation in Design Science Research. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7286 LNCS*, 423–438. https://doi.org/10.1007/978-3-642-29863-9_31

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. *Information Systems Research*, *3*(1), 36–59. https://doi.org/10.1287/ISRE.3.1.36

Weir, C. B., & Jan, A. (2021). BMI Classification Percentile And Cut Off Points. *StatPearls*. https://www.ncbi.nlm.nih.gov/books/NBK541070/

Wildschut, A. (2010). Doctors in the public service: too few for too many. *HSRC Review*, *8*, 12–15. http://www.hsrc.ac.za/en/review/november-/public-service-doctors

Williams, C. (2007). Research Methods. *Journal of Business & Economics Research (JBER)*, *5*(3), 65. https://doi.org/10.19030/JBER.V5I3.2532

World Health Organisation. (2016). *WHO: Stillbirths*. World Health Organization. https://www.who.int/maternal_child_adolescent/epidemiology/stillbirth/en/

Wu, Y., Moylan, E., Inman, H., & Graf, C. (2019). Paving the Way to Open Data. *Data Intelligence*, *1*(4), 368–380. https://doi.org/10.1162/DINT_A_00021

Zhang, S., Zhang, C., & Yang, Q. (2016). *Data preparation for data mining*. https://doi.org/10.1080/713827180

Appendices

APPENDIX A: Ethical Clearance Letter - H21-SCI-CSS-005

NELSON MANDELA
UNIVERSITY

PO Box 77000, Nelson Mandela University, Port Elizabeth, 6031, South Africa www.mandela.ac.za

Vice-Chairperson: Research Ethics Committee (Human)
Tel: +27 (0)41 504 2630
Zoleka.Soji@mandela.ac.za

NHREC registration nr: REC-042508-025

Ref: [H21-SCI-CSS-005 / Amendment]

24 January 2022

Prof J Wesson
Faculty: Science

Dear Prof Wesson

DEVELOPING A MACHINE LEARNING MODEL TO PREDICT THE DEVELOPMENT OF HYPERTENSION DURING PREGNANCY

PRP: Prof J Wesson
PI: Mr N Smith

The request for an amendment to the above-entitled study was considered for expedited approval by the Research Ethics Committee (Human) (21 January 2022). We take pleasure in informing you that the Research Ethics Committee (Human) approved the amendment. The study is classified as medium risk. The ethics number remains **[H21-SCI-CSS-005]**; approval is subject to the following conditions:

1. The immediate completion and return of the attached acknowledgement to Imtiaz.Khan@mandela.ac.za.
2. The submission of an annual progress report by the PRP on the data collection activities of the study (form RECH-004 available on Research Ethics Committee (Human) portal) by 15 November this year for studies approved/extended in the period October of the previous year up to and including September of this year, or 15 November next year for studies approved/extended after September this year.
3. In the event of a requirement to extend the period of data collection (i.e. for a period in excess of 1 calendar year from date of approval), completion of an extension request is required (form RECH-005 available on Research Ethics Committee (Human) portal)
4. In the event of any changes made to the study (excluding extension of the study), RECH will have to approve such amendments and completion of an amendments form is required PRIOR to implementation (form RECH-006 available on Research Ethics Committee (Human) portal).
5. Immediate submission (and possible discontinuation of the study in the case of serious events) of the relevant report to RECH (form RECH-007 available on Research Ethics Committee (Human) portal) in the event of any unanticipated problems, serious incidents or adverse events observed during the course of the study.
6. Immediate submission of a Study Termination Report to RECH (form RECH-008 available on Research Ethics Committee (Human) portal) upon expected or unexpected closure/termination of study.
7. Immediate submission of a Study Exception Report of RECH (form RECH-009 available on Research Ethics Committee (Human) portal) in the event of any study deviations, violations and/or exceptions.
8. Acknowledgement that the study could be subjected to passive and/or active monitoring without prior notice at the discretion of Research Ethics Committee (Human).

APPENDIX B: Demographic and baseline clinical features of women who had stillbirths in Madhi et al.'s (2019) study

	Total (n=350)	Complete sampling (n=294)	Incomplete sampling (n=56)	p value*
Age, years	27 (23–33)	27 (23–33)	27 (25–31)	0·87
≥35	59/344 (17%)	52/289 (18%)	7/55 (13%)	0·44
<35	285/344 (83%)	237/289 (82%)	48/55 (87%)	..
MUAC, cm	29·4 (4·5)	29·4 (4·6)	29·6 (4·2)	0·79
>33 (obese)	51/291 (18%)	42/248 (17%)	9/43 (21%)	0·52
<23 (malnourished)	11/291 (4%)	10/248 (4%)	1/43 (2%)	>0·99
Parity (before current pregnancy)	1 (1–2)	1 (1–2)	1 (1–2)	0·35
0	44/253 (17%)	40/212 (19%)	4/41 (10%)	0·12
1–4	204/253 (81%)	169/212 (80%)	35/41 (85%)	..
≥5	5/253 (2%)	3/212 (1%)	2/41 (5%)	..
At least one antenatal care visit	322/350 (92%)	271/294 (92%)	51/56 (91%)	0·79
Multiple gestation (twins)	14/350 (4%)	12/294 (4%)	2/56 (4%)	>0·99
Previous stillbirth	26/350 (7%)	19/294 (6%)	7/56 (13%)	0·091
Timing of stillbirth	..	..	..	0·45
Antepartum	227/350 (65%)	188/294 (64%)†	39/56 (70%)	..
Intrapartum	123/350 (35%)	106/294 (36%)‡	17/56 (30%)	..
Diagnosis of stillbirth	..	..	..	>0·99
After delivery	54/344 (16%)	46/289 (16%)	8/55 (15%)	..
Predelivery ultrasound	290/344 (84%)	243/289 (84%)	47/55 (85%)	..
Mode of delivery	..	..	..	0·87
Vaginal	259/350 (74%)	218/294 (74%)	41/56 (73%)	..
Caesarean section	91/350 (26%)	76/294 (26%)	15/56 (27%)	..
HIV infected	94/350 (27%)	82/294 (28%)	12/56 (21%)	0·41
Receiving antiretroviral therapy for HIV	82/94 (87%)	75/82 (91%)	7/12 (58%)	0·0070
Duration of antiretroviral therapy, months§	2 (1–4)	2 (1–4)	3 (3–4)	0·37
Tobacco smoker¶	6/195 (3%)	4/163 (2%)	2/32 (6%)	0·26
Alcohol use during pregnancy¶	9/190 (5%)	6/159 (4%)	3/31 (10%)	0·17
Maternal syphilis				
Tested	344/350 (98%)	292/294 (99%)	52/56 (93%)	0·0069
Positive	6/344 (2%)	6/292 (2%)	0/52	0·60
Pregnancy-induced hypertension	111/350 (32%)	93/294 (32%)	18/56 (32%)	>0·99
Chronic hypertensive disorder	22/350 (6%)	18/294 (6%)	4/56 (7%)	0·76
Pregnancy-associated diabetes	8/350 (2%)	7/294 (2%)	1/56 (2%)	>0·99
Other chronic medical conditions‖	2/350 (1%)	2/294 (1%)	0/56	>0·99
Intrapartum complications				
Cord prolapse	4/350 (1%)	4/294 (1%)	0/56	>0·99
Prolonged labour**	31/129 (24%)	25/107 (23%)	6/22 (27%)	0·78
Fetal distress	34/350 (10%)	30/294 (10%)	4/56 (7%)	0·63
Placental abnormalities (clinically diagnosed)				
Placenta praevia	2/350 (1%)	1/294 (<1%)	1/56 (2%)	0·30
Placental abruption††	67/350 (19%)	62/294 (21%)	5/56 (9%)	0·040
Uterine rupture	4/350 (1%)	4/294 (1%)	0/56	>0·99
Evidence of maternal infection during pregnancy or labour	27/350 (8%)	23/294 (8%)	4/56 (7%)	>0·99
Intrapartum temperature ≥38°C	4/350 (1%)	4/294 (1%)	0/56	>0·99
Clinically diagnosed chorioamnionitis	3/350 (1%)	2/294 (1%)	1/56 (2%)	0·41
Urinary tract infection	14/350 (4%)	11/294 (4%)	3/56 (5%)	0·48
Malodorous or purulent vaginal discharge	8/350 (2%)	7/294 (2%)	1/56 (2%)	>0·99
Uterine tenderness	1/350 (<1%)	1/294 (<1%)	0/56	>0·99

Data are n/N (%), mean (SD), or median (IQR). MUAC=mid-upper arm circumference. *Complete versus incomplete sampling groups. †Delivered 190 stillbirths. ‡Delivered 108 stillbirths. §From time of initiation to time of birth in HIV-infected women. ¶Quantity not ascertainable from medical records. ‖Includes one mother with cardiac abnormality and one with thyroid dysfunction. **Defined as ≥20 h in nulliparous and ≥14 h in multiparous women. ††Overall 38 (13%) of 290 stillbirths diagnosed predelivery on ultrasound, including 37 (15%) of 243 with complete sampling, and 1 (2%) of 47 with incomplete sampling (p=0·014).

APPENDIX C: Paper Published in the Southern African Telecommunications, Network and Communication (SATNAC) 2021 conference

Using Machine Learning to Predict Preeclampsia in Pregnant Women in South Africa

Nathan Smith[1], Prof Janet Wesson[2]

Department of Computing Sciences, Nelson Mandela University, Port Elizabeth, South Africa

[1]Nathan.Smith@Mandela.ac.za

[2]Janet.Wesson@Mandela.ac.za

Abstract— **Preeclampsia is one of the leading causes of maternal mortality in South Africa and the world. Due to South Africa's poor antenatal care, the prediction of pregnant women at risk of developing preeclampsia is vital for allocating scarce resources. This paper proposes a machine learning model to predict preeclampsia using existing antenatal care datasets from South African hospitals. To the researcher's knowledge, this proposed model will be the first machine learning model for predicting preeclampsia using a South African dataset. A review of the literature and existing systems was performed to identify the eight risk factors. These risk factors are systolic blood pressure, diastolic blood pressure, maternal age, body mass index, diabetes status, hypertension history, nulliparity, and maternal disease. The proposed model will use supervised learning and will be evaluated using five evaluation metrics, namely classification accuracy, confusion matrix, logarithmic loss, area under curve, and F-Score.**

Keywords— **Preeclampsia, Pregnancy, Predictions, Forecasting, Machine Learning, Supervised Learning, Proposed Model**

INTRODUCTION

Approximately 800 women die every day worldwide from preventable causes due to pregnancy and childbirth. Ninety-four per cent of these maternal deaths occur in low- to middle-income countries, with Sub-Saharan Africa accounting for two-thirds of these deaths (*Maternal Mortality*, n.d.). Hypertensive disorders of pregnancy, such as preeclampsia, are the second leading causes of maternal death in South Africa (Nathan et al., 2018). Preeclampsia, usually diagnosed after 20 weeks' gestation, is the new onset of hypertension. It is a disorder characterised by high blood pressure, usually above 140/90, and the presence of proteinuria (Roberts & Gammill, 2005). The mortality of pregnant women with preeclampsia is accompanied by increased foetal mortality and disability rates (Rosser & Katz, 2013), (Savaj & Vaziri, 2012). The maternal complications associated with preeclampsia are acute kidney disease and placental abruption. In severe circumstances, preeclampsia can lead to low platelet count, elevated liver enzymes, haemolysis and eclamptic seizure.

Preeclampsia can be categorised into two groups, early- and late-onset. Early-onset preeclampsia develops before 34 weeks' gestation, and late-onset preeclampsia develops after 34 weeks' gestation. In South Africa, preeclampsia incidence is much higher than reported in other low- and middle-income studies. These statistics could be explained by South African women lacking access to antenatal and postnatal care, resulting in delays in seeking care (Nathan et al., 2018). If they can find a clinic, many pregnant women only attend their first session or attend it very late in their pregnancy (Ebonwu, Mumbauer, Uys, Wainberg, & Medina-Marino, 2018). South Africa has been fighting this ongoing battle with access to antenatal care and preeclampsia for decades, despite the availability of free antenatal care throughout the country (Myer & Harrison, 2003), (Ebonwu, Mumbauer, Uys, Wainberg, & Medina-Marino, 2018).

The importance of antenatal care for pregnant women with hypertensive disorders is echoed in a meta-analysis of the prevalence of hypertensive disorders of pregnancy, and pregnancy outcomes in Sub-Saharan Africa. The meta-analysis recommends future research in strategies to predict those women at greater risk of hypertensive disorders. Since there

are no preventative measures for preeclampsia, clinics focus on early detection and surveillance. Early detection will allow for the prioritisation of preeclampsia patients, especially in low resource settings. Due to the large amounts of data the health sector produces, machine learning offers a useful way to extract meaning from this data (Bhardwaj, Nambiar, & Dutta, 2017). The health sector has already seen the positive impact that machine learning has had on the accuracy of prediction in healthcare (Bottaci et al., 1997; Frizzell et al., 2017). This paper proposes a model that will predict if a pregnant woman is at risk of developing preeclampsia.

Related Work

Jhee et al. [11] developed a model to predict late-onset preeclampsia using machine learning-based methods. They included data from 11,006 women from the Yonsei University Healthcare Centre in China between 2005 and 2017. The preeclampsia development rate was 4.7% (n=474). The data included were: age, blood pressure, height, weight, gestational age, medical history, and biochemical laboratory data. The repeated-measures data, such as blood pressure, weight and laboratory data were delineated through pattern recognition and cluster analysis. Their model included a data split of 70% training and 30% testing, and two outcome categories, namely preeclampsia and no preeclampsia. Model development was programmed in R. The following six methods were used: logistic regression, decision tree model, naïve Bayes classification, support vector machine, random forest algorithm, and stochastic gradient boosting method. Jhee et al. [11] made use of the R programming language for all their models.

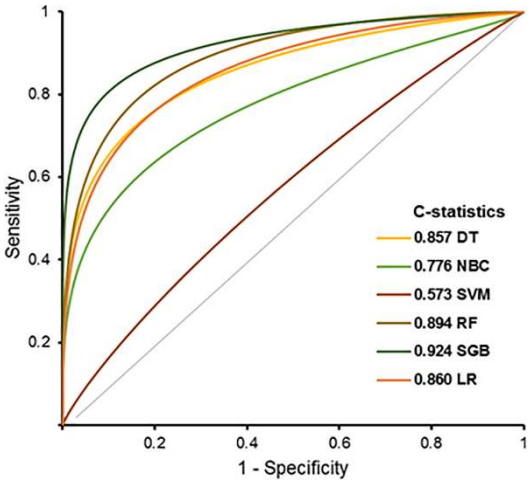

Figure 1: Jhee's Model's calibration plot with respective C-statistics (Jhee et al., 2019)

When the prediction performances were compared among the prediction models, the stochastic gradient boosting (SGB) model had the best performance for predicting preeclampsia. The overall accuracy of the SGB model was 0.973, the false-positive rate was 0.009, and the detection rate reached 0.771. Jhee et al. [11] highlighted the use of mean values when dealing with prediction models in pregnancy. They agree on the importance of taking fluctuation variability into account, and this is why they incorporated repeated measured values, and included the changing patterns as an analysable factor. These recommendations support the idea of incorporating smart devices such as smartwatches to automatically monitor fluctuation. Although Jhee et al. [11] could effectively predict

Pattern recognition and cluster analysis were used to determine the influence of each variable on prediction. Among the assessed variables, the 14 most influential factors were included in their prediction model. The most influential risk factor was systolic blood pressure. The C-statistic, a global measure of model discrimination, for all models used by Jhee et al [11] accompanied with their respective calibration plots are shown in Fig. 1.

late-onset preeclampsia, they faced several limitations. Many women only started the antenatal evaluation program early in the second trimester, resulting in unusable first trimester data. Acquiring large amounts of data will allow this study to overcome this limitation. When the data processing stage has been reached, all participants that started the antenatal care program late can be removed without effecting the models performance. Jhee et al. [11] used a dataset containing 11,006 women, although this is a large sample and even larger sample may provide a more accurate model.

One simple prediction system currently used in South African clinics is the basic antenatal care approach (Bantenatal care). The Bantenatal care approach has been simplified to allow all postnatal care midwives to provide antenatal services to reduce maternal and perinatal deaths, while improving maternal health (Sylvia Patience Ngxongo, 2019). The Bantenatal care approach allows for better decision-making at the primary healthcare level. It achieves this by identifying high-risk pregnancies at an early stage. Although it is a very simple paper-based approach to predicting high-risk pregnancy, it has shown to be very effective.

In an article by Hofmeyr and Mentrop [15] in the South African Medical Journal, a case study is discussed that best describes the impact of poor antenatal care in South Africa, as follows: "A 22-year-old nulliparous woman was seen antenatally at 26 weeks' and 35 weeks' gestation and was well. Her blood pressure was 110/60 mmHg, urine tests were normal, and foetal movements were felt. Her next visit was booked for six weeks' later. After 23 days, she presented at 08h30 in labour with severe pulmonary oedema. Her blood pressure was 189/93 mmHg, and oxygen saturation was 70% on room air and 85% on 40% oxygen by mask. After stabilisation, a caesarean section was performed for foetal distress under general anaesthesia. A male baby was delivered

at 09h50, with an Apgar score of 6/10 at 5 minutes. During the closure of the uterus, the mother had a cardiac arrest. Resuscitation was carried out, and she was transferred to the intensive care unit. Her condition deteriorated despite intensive care, and she died at 15h40. The most crucial avoidable factor was the long interval between routine Bantenatal care antenatal visits. Under the traditional antenatal model, she would have been seen two weeks after her visit at 35 weeks, and early preeclampsia would likely have been diagnosed and managed with delivery before she progressed to severe preeclampsia with pulmonary oedema" (Hofmeyr & Mentrop, 2015). This case study illustrates the importance of proper antenatal care and proper management of resources.

Another study consulted was a systematic literature review and large meta-analysis (Bartsch et al., 2016). The study identified publications investigating the association between preeclampsia and at least one risk factor in a previous pregnancy or the current pregnancy. Pubmed and Embase were searched for in English papers available from 2000 to June 2016. The pooled preeclampsia event rate for each risk factor was calculated using an arcsine transformation, followed by the derivation of a pooled relative risk (RR_{pooled}), and 95 per cent confidence intervals for each variable. Using the pooled relative risks, the population attributable fraction for each risk factor was calculated using the following formula:

$$PAF=[P_{epooled}\ (RR_{pooled}-1)]/[P_{epooled}(RR_{pooled}-1)+1]$$

Where $P_{epooled}$ is the number of women with a given risk factor in each study, divided by the total number of women in the study. Table I contains the pooled unadjusted risk for each risk factor.

TABLE

RISK OF PREECLAMPSIA DETERMINED BY 16 WEEKS' GESTATION (Bartsch et al., 2016)

Risk Factor	**No of women/No of studies**	**Pooled unadjusted risk (95% CI)**
Prior IUGR	55,542/1	1.4(0.6 to 3.0)
SLE	2,413,908/2	2.5 (1.0 to 6.3)
Nulliparity	2,975,158/25	2.1 (1.9 to 2.4)
Maternal Age > 35	5,244,543/22	1.2 (1.1 to 1.3)
Maternal age > 40	4,260,202/16	1.5 (1.2 to 2.0)
Prior stillbirth	63,814/2	2.4 (1.7 to 3.4)
Chronic kidney disease	966,505/5	1.8 (1.5 to 2.1)
ART	1,463,529/20	1.8 (1.6 to 2.1)
Pre-pregnancy BMI>25	3,644,747/38	2.1 (2.0 to 2.2)
Pre-pregnancy BMI>30	5,921,559/40	2.8 (2.6 to 3.1)
Multifoetal pregnancy	7,309,227/8	2.9 (2.6 to 3.1)
Prior placental abruption	291,134/3	2.0 (1.4 to 2.7)
Pregestational diabetes	2,553,117/19	3.7 (3.1 to 4.3)
Prior preeclampsia	3,720,885/20	8.4 (7.1 to 9.9)
Chronic hypertension	6,589,661/20	5.1 (4.0 to 6.5)
aPL	220,156/3	2.8 (1.8 to 4.3)

*IUGR=intrauterine growth restriction; SLE=systemic lupus erythematosus; ART=assisted reproductive technology; BMI=body mass index; aPL=antiphospholipid antibody syndrome

Prior preeclampsia had the most significant pooled relative risk, followed by chronic hypertension. Nulliparity had the greatest population attributable fraction for preeclampsia with pre-pregnancy BMI > 25 and prior preeclampsia in the second and third most. The study concluded by stating that all identified risk factors attributed to a heightened risk of preeclampsia. The most prominent risk factors were chronic hypertension, pre-gestational diabetes, prior preeclampsia, and BMI > 30.

The quality and quantity of data are some of the most influential factors regarding accurate predictions in machine learning models. While performing the literature review for this research, a clear gap in the accessibility of antenatal data for research purposes was identified. The Global Pregnancy Collaboration (CoLab) is a company currently striving to promote data accessibility and standardisation among research in preeclampsia (Ebonwu, Mumbauer, Uys, Wainberg, & Medina-Marino, 2018). Their goal is to create a large and powerful dataset by merging datasets from smaller studies worldwide. Standardised data is another key selling point for the COLLECT dataset. The time required to merge smaller datasets is significant, primarily due to data collection projects producing vastly different data sets. All projects currently using the COLLECT dataset are still in their data collection phase. Thus, this data cannot be used for this paper's proposed model. However, it is recommended that future research is aimed at using the COLLECT dataset for South African research in preeclampsia (J. E. Myers, Myatt, Roberts, & Redman, 2019).

PREECLAMPSIA IN PREGNANT WOMEN

Preeclampsia is a disorder associated with pregnancy, affecting 4.6% (95% confidence interval (CI), 2.7–8.2) of all pregnancies. It remains a leading cause of maternal and perinatal morbidity and mortality worldwide (Abalos, Cuesta, Grosso, Chou, & Say, 2013). In the guidelines published by the National Institute for Health and Care Excellence (NICE) in 2019, a woman is classified at high risk of preeclampsia if there is a history of hypertensive disease during a previous pregnancy or a maternal disease, including chronic kidney disease, autoimmune diseases, diabetes, or chronic hypertension. Women are at moderate risk if they are (M. A. Brown et al., 2018)(NICE, 2019):

- Nulliparous

- ≥ 40 years of age
- Body mass index (BMI) ≥ 35 kg/m
- Family history of preeclampsia
- A multi-foetal pregnancy
- Pregnancy interval > 10 years

The presence of one high-risk factor, or two or more moderate risk factors, is used to help guide aspirin prophylaxis, which helps to reduce the risk of preeclampsia if administered before 16 weeks' gestation (Askie, Duley, Henderson-Smart, & Stewart, 2007)(Bujold et al., 2010).

After consulting relevant literature, the following risk factors for preeclampsia were identified [11] – [19]:

- Systolic Blood Pressure
- Diastolic Blood Pressure
- Maternal Age
- BMI
- Diabetes Status
- Hypertension History
- Nulliparity
- Maternal Disease

However, pre-collected datasets may contain more or less features that were not explicitly highlighted. Therefore, these risk factors will be used as a guideline.

Machine Learning for Prediction of Preeclampsia

The healthcare sector is always striving to achieve the Triple Aim, which improves outcomes, enhances patients' experience, and reduces healthcare costs to the public. Predictive modelling for real-time decision making is playing an essential role in achieving the Triple Aim. A clinical risk prediction model is defined as a model that combines several characteristics to predict the risk of disease presence and outcome occurrence in individuals (Alonzo, 2009). Risk-prediction models as decision-making tools have long played an essential role in clinical practice. However, as times change, so do risk-prediction models. Two significant changes in risk prediction have been centred around advances in computer science. Firstly, we now live in a data-centric world where almost every aspect of our lives is digitally stored in some format. Secondly, there has been a sudden increase in the use of machine learning models to make predictions. The significant increase in data has led to an increasing need to analyse and interpret it (Jordan & Mitchell, 2015). Machine learning models allow for large amounts of data to be analysed and interpreted in a fraction of the time it would take a human. Thus, they save time and are immune to 'human error'. Machine learning is extensive and can include simple models such as the Bantenatal care model, or highly complex models such as Tesla's complex deep neural network for their autopilot (*Autopilot AI | Tesla*, n.d.).

Proposed Model

Machine learning models follow an iterative approach where a threshold is selected and, if not reached, will result in the model being adapted (Figure 2).

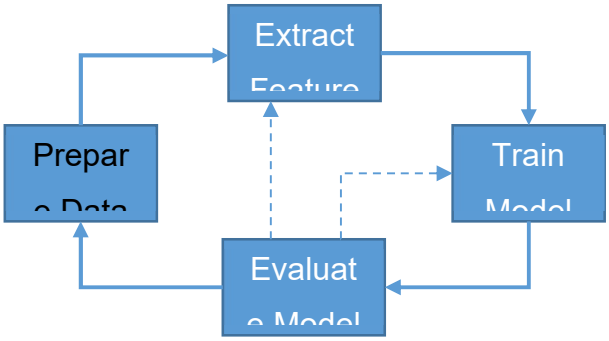

Figure II: Iterative Machine Learning Process (*"Iterative Learning-InsideAIML,"* n.d.)

There are seven general steps to take when developing a machine learning model (Lantz, 2013), (Swamynathan, 2017). These steps are:

1. Acquiring data
2. Preparing data
3. Choosing a suitable model
4. Training the model
5. Evaluating the model
6. Tuning the parameters
7. Making predictions

All seven steps will be described in detail in the following paragraph. The starting point for any machine learning model is the acquisition of data.

Data Acquisition

Clean and descriptive data is an essential aspect of a machine learning model, as without it, there is no way of making an accurate prediction. There are two main methods for acquiring data, either by performing the data collection or using an existing data set. Initially, the proposed model was going to make use of data collected from smartwatches of patients in South African clinics. However, due to the Covid pandemic, cost, logistics and time constraints, this proved unfeasible. Thus, an existing antenatal care dataset was needed.

The lack of customisation is a big disadvantage of using an existing dataset, and must be considered when selecting which dataset to use. Some risk factors that are crucial for making an accurate prediction may not be included in the dataset. Therefore, consulting many datasets to find which one suits the proposed model and contains most of the required features is very important. The researcher investigated many public dataset repositories for antenatal care or preeclampsia datasets. Unfortunately, no suitable public dataset could be found. Thus, another avenue had to be explored, namely datasets directly

from health institutions. Many hospitals in South Africa store their data and allow researchers to apply for access to the anonymised datasets. However, before applying for access to the data, ethical clearance from the Nelson Mandela University needs to be acquired. This ethical clearance is currently underway, with the application awaiting review. Thereafter, applications for access to the hospitals' antenatal datasets will be sent.

Data Preparation

Data preparation follows data acquisition, and the emphasis is on data quality. Data preparation is the transformation of raw data into a form suited for the machine learning model. Before handling the data, it is crucial to visualise it to highlight any relationships or data imbalances. The data preparation techniques used will depend on the specifics of the data. Generally, the first task is data cleaning. Data cleaning is the process of identifying and fixing systematic problems. General data operations performed in the data cleaning are:

- Use statistics to identify outliers by defining a 'normal' value;
- Identify and remove duplicates;
- Identify null values and replace them using statistics.

The next step is feature selection. Feature selection refers to the selection of a subset of input features that most affect the output. There is no 'best feature selection method'. Each specific problem will require experimentation of methods. The three most common feature selection methods for supervised learning problems are:

- *Wrapper:* The search for the best performing subset of features;
- *Filter:* The selection of a subset of features based on their relationship with the target variable
- *Intrinsic:* Algorithms that perform automatic feature selection during training, such as decision trees.

Once the optimal features have been selected, data transformation needs to begin. Data transformation is used to change the type of distribution of data. There are many different techniques, and only once the data has been thoroughly analysed will the suitable methods be identified. The final step in data preparation is feature engineering, which is the creation of new input variables from the existing available data. Feature engineering is used to add broader context to a single observation or to simplify a complex variable. This can be done by adding a Boolean flag, adding a summary statistic, or breaking down existing variables. An example of feature engineering for the proposed model would be adding a Boolean field that would be derived from the maternal age. Adding this field could potentially decrease the processing time of the model.

Once all these steps are performed the data will be randomly split into two sets, one being a training set equalling 70 per cent of the complete data set, and the remaining 30 per cent will be used as a testing set. Thus, the data will be ready to train and evaluate the model.

Model Selection

The three main types of machine learning models are supervised, unsupervised and semi-supervised learning. Supervised learning, which this paper's proposed model will use, requires all data to be labelled. The model then learns a mapping between input and output variables, and applies it to unseen data to predict the outputs (Cunningham, Cord, & Delany, 2008). Supervised learning can be broken down into two types of problems, namely classification and regression. A classification algorithm is used when you have a categorical outcome, such as true or false. A regression algorithm is used when you have a real value output, such as weight. Therefore, since the outcome for the proposed model will be either at risk of hypertension or not at risk, classification will be the selected algorithm. Women who developed preeclampsia will be categorised into the preeclampsia group, and those who did not, would be placed in the no preeclampsia group. Some of the fundamental algorithms that will be tested are:

- Support Vector Machine
- Naïve Bayes
- Decision Tree
- Artificial Neural Network

Model Training

Once the model is selected, the data will be used to train the model. Training a model requires the data to be supplied to the model to incrementally improve the model's ability to make predictions. Python will be the selected language due to the first author's familiarity with it, and that it is the most used language for programming machine learning models. The SciKit library will be used for creating and training the model (Pedregosa et al., 2011).

Model Evaluation

After the model has been trained with the training dataset, the model will have to be evaluated. The evaluation phase uses the testing dataset, which was set aside in the data processing phase. The following metrics are used to quantify the classification model's performance:

- *Classification accuracy* – number of predictions made as a ratio of all predictions made;
- *Confusion matrix* – A detailed breakdown of correct and incorrect classifications for each class;
- *Logarithmic Loss* – Performance of the model where the prediction probability diverges from the actual label. The goal is to minimise this value;
- *Area under curve* – A measure of the ability of a binary classifier to discriminate between positive and negative classes;
- *F-Measure (F-Score)* – Considers both the recall and the precision to compute a score for the model.

The overall performance of the proposed model can be evaluated using all the metrics.

Parameter Tuning

If the model's accuracy is not adequate, parameter tuning will be needed. It must be noted that before parameter tuning is performed, an attempt to improve the quality of data being input into the model must be performed. If the data cannot be improved, two methods for tuning the parameters will be used: a grid search and a random search. A grid search evaluates a model for each combination of parameters specified in a grid, whereas a random search samples each set of parameters from a distribution over a possible parameter value. Once the tuning is complete, the best configuration will be reported, and the necessary steps can be taken.

Performing Predictions

Finally, if the model's performance is satisfactory, the prediction model can be tested in a real-world scenario. When comparing studies with similar models, Jhee et al. achieved an overall accuracy on their SGB model of 0.973, a false positive rate of 0.009, and a detection rate of 0.771. This paper's proposed model will aim to perform equally or better than Jhee et al.'s SGB model.

Discussion and Future Work

After performing the literature review and applying for ethics, the next step in developing the model is to apply for a usable data set. Thereafter, data preparation will begin (Section V-*B*). Figure 3 visualises the current status of the development of the proposed model.

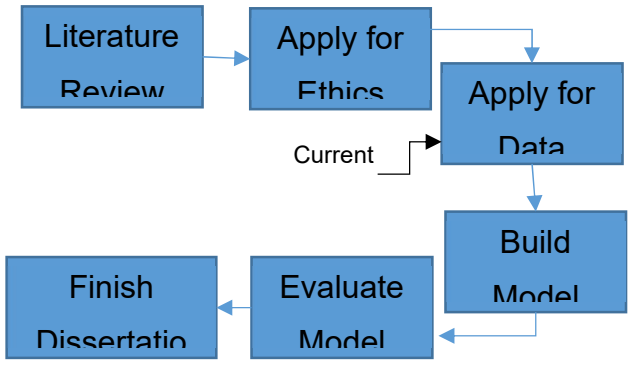

Figure I: Visualisation of the proposed model development status

Once data preparation begins, the goal will be to use a South African hospital's antenatal dataset as the proposed model's primary data source. Once implemented, the model will be evaluated using all the methods outlined in Section V. The final step would be to implement the model in a South African clinic or hospital. If the model does not perform within acceptable thresholds, independent data collection will need to be considered. If data collection is needed, the COLLECT dataset and standards will be used.

The contribution of this paper is a proposed model for the prediction of preeclampsia using machine learning techniques. To the researcher's knowledge, this proposed model is the first of its kind in South Africa and the first to make use of a South African antenatal data set. Thus, the performance of different machine learning algorithms on South African antenatal data can be evaluated.

Conclusion

Preeclampsia continues to affect pregnant South African women and their families. This paper proposes a model, which is the first of its kind in South Africa, for accurately predicting women at risk of developing preeclampsia, allowing clinics to assign their limited resources to women in critical need. The proposed model is planned to be the first preeclampsia prediction model that makes use of a South African Hospital's dataset. This paper plans on improving existing preeclampsia prediction models by increasing the sample size and exploring different prediction techniques. Future work will focus on the accessibility and standardisation of antenatal care data for future research into preeclampsia.

Acknowledgements

This work was supported by the Telkom/NMU Centre of Excellence and the Nelson Mandela University Research Centre.

References

Abalos, E., Cuesta, C., Grosso, A. L., Chou, D., & Say, L. (2013). Global and regional estimates of preeclampsia and eclampsia: A systematic review. In *European Journal of Obstetrics and Gynecology and Reproductive Biology* (Vol. 170, Issue 1, pp. 1–7). Elsevier Ireland Ltd. https://doi.org/10.1016/j.ejogrb.2013.05.005

Afrose, S., Song, W., Nemeroff, C. B., Lu, C., Yao, D., & Author, C. (2021). Overcoming Underrepresentation in Clinical Datasets for Accurate Subpopulation-specific Prognosis. *MedRxiv*, 2021.03.26.21254401. https://doi.org/10.1101/2021.03.26.21254401

Airoldi, J., & Weinstein, L. (2007). Clinical significance of proteinuria in pregnancy. *Obstetrical and Gynecological Survey*, *62*(2), 117–124. https://doi.org/10.1097/01.OGX.0000253301.55009.AC

Akbari, S., Khodadadi, B., Ahmadi, S. A. Y., Abbaszadeh, S., & Shahsavar, F. (2018). Association of vitamin D level and vitamin D deficiency with risk of preeclampsia: A systematic review and updated meta-analysis. *Taiwanese Journal of Obstetrics & Gynecology*, *57*(2), 241–247. https://doi.org/10.1016/J.TJOG.2018.02.013

Alonzo, T. A. (2009). Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating: By Ewout W. Steyerberg. *American Journal of Epidemiology*, *170*(4), 2. https://doi.org/10.1093/aje/kwp129

Aminu, M., Unkels, R., Mdegela, M., Utz, B., Adaji, S., & van den Broek, N. (2014). Causes of and factors associated with

stillbirth in low- and middle-income countries: a systematic literature review. *BJOG : An International Journal of Obstetrics and Gynaecology*, *121*, 141–153. https://doi.org/10.1111/1471-0528.12995

An, C., Lim, H., Kim, D. W., Chang, J. H., Choi, Y. J., & Kim, S. W. (2020). Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. *Scientific Reports 2020 10:1*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-75767-2

Ananth, C. V., Keyes, K. M., & Wapner, R. J. (2013a). Pre-eclampsia rates in the United States, 1980-2010: age-period-cohort analysis. *BMJ (Clinical Research Ed.)*, *347*. https://doi.org/10.1136/BMJ.F6564

Ananth, C. V., Keyes, K. M., & Wapner, R. J. (2013b). Pre-eclampsia rates in the United States, 1980-2010: Age-period-cohort analysis. *BMJ (Online)*, *347*. https://doi.org/10.1136/bmj.f6564

Askie, L. M., Duley, L., Henderson-Smart, D. J., & Stewart, L. A. (2007). Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data. *Lancet*, *369*(9575), 1791–1798. https://doi.org/10.1016/S0140-6736(07)60712-0

Autopilot AI | Tesla. (n.d.). Retrieved June 17, 2021, from https://www.tesla.com/autopilotAI

Ayo-Yusuf, L. (2015). *South Africa needs a new way to address the doctor shortage*. https://theconversation.com/south-africa-needs-a-new-way-to-address-the-doctor-shortage-41136

Bartsch, E., Medcalf, K. E., Park, A. L., Ray, J. G., Al-Rubaie, Z. T. A., Askie, L. M., Berger, H., Blake, J., Graves, L., Kingdom, J. C., Lebovic, G., Lord, S. J., Maguire, J. L., Mamdani, M. M., Meloche, J., Urquia, M. L., & Van Wagner, V. (2016). Clinical risk factors for pre-eclampsia determined in early pregnancy: Systematic review and meta-analysis of large cohort studies. *The BMJ*, *353*. https://doi.org/10.1136/bmj.i1753

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design Science Research Contributions: Finding a Balance between Artifact and Theory. *Journal of the Association for Information Systems*, *19*(5), 358–376. https://doi.org/10.17705/1jais.00495

Battarbee, A. N., Sinkey, R. G., Harper, L. M., Oparil, S., & Tita, A. T. N. (2020). Chronic Hypertension in Pregnancy. *American Journal of Obstetrics and Gynecology*. https://doi.org/10.1016/j.ajog.2019.11.1243

Berwick, D. M., Nolan, T. W., & Whittington, J. (2017). The Triple Aim: Care, Health, And Cost. *Https://Doi.Org/10.1377/Hlthaff.27.3.759*, *27*(3), 759–769. https://doi.org/10.1377/HLTHAFF.27.3.759

Bezerra, P. C. F. M., Leão, M. D., Queiroz, J. W., Melo, E. M. D., Pereira, F. V. M., Nóbrega, M. H., Jeronimo, A. K., Ferreira, L. C., Jerônimo, S. M. B., & De Araújo, A. C. P. F. (2010). Family history of hypertension as an important risk factor for the development of severe preeclampsia. *Acta Obstetricia et Gynecologica Scandinavica*, *89*(5), 612–617. https://doi.org/10.3109/00016341003623720

Bhardwaj, R., Nambiar, A. R., & Dutta, D. (2017). A Study of Machine Learning in Healthcare. *Proceedings - International Computer Software and Applications Conference*, *2*, 236–241. https://doi.org/10.1109/COMPSAC.2017.164

Births and Deaths Registration Act 51 of 1992, (1992). https://www.gov.za/sites/default/files/gcis_document/201409/a511992.pdf

Bottaci, L., Drew, P. J., Hartley, J. E., Hadfield, M. B., Farouk, R., Lee, P. W. R., MacIntyre, I. M. C., Duthie, G. S., & Monson, J. R. T. (1997). Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *Lancet*, *350*(9076), 469–472. https://doi.org/10.1016/S0140-6736(96)11196-X

Boulle, A., Heekes, A., Tiffin, N., Smith, M., Mutemaringa, T., Zinyakatira, N., Phelanyane, F., Pienaar, C., Buddiga, K., Coetzee, E., van Rooyen, R., Dyers, R., Fredericks, N., Loff, A., Shand, L., Moodley, M., de Vega, I., & Vallabhjee, K. (2019). Data Centre Profile: The Provincial Health Data Centre of the Western Cape Province, South Africa. *International Journal of Population Data Science*, *4*(2). https://doi.org/10.23889/IJPDS.V4I2.1143

Brantsæter, A. L., Haugen, M., Samuelsen, S. O., Torjusen, H., Trogstad, L., Alexander, J., Magnus, P., & Meltzer, H. M. (2009). A Dietary Pattern Characterized by High Intake of Vegetables, Fruits, and Vegetable Oils Is Associated with Reduced Risk of Preeclampsia in Nulliparous Pregnant Norwegian Women. *The Journal of Nutrition*, *139*(6), 1162. https://doi.org/10.3945/JN.109.104968

Bresnick, J. (2017). *Google's Machine Learning, Imaging Analytics Flag Breast Cancer*. https://healthitanalytics.com/news/googles-machine-learning-imaging-analytics-flag-breast-cancer

Brown, M. A., Magee, L. A., Kenny, L. C., Karumanchi, S. A., McCarthy, F. P., Saito, S., Hall, D. R., Warren, C. E., Adoyi, G., & Ishaku, S. (2018). Hypertensive disorders of pregnancy: ISSHP classification, diagnosis, and management recommendations for international practice. In *Hypertension* (Vol. 72, Issue 1, pp. 24–43). Lippincott Williams and Wilkins. https://doi.org/10.1161/HYPERTENSIONAHA.117.10803

Brown, P. (2018). *Noncommunicable diseases: Risk factors*. https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/ncd-risk-factors

Brownlee, J. (2020). *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python* (v1.2). Machine Learning Mastery, 2020. https://books.google.co.za/books?hl=en&lr=&id=uAPuDwAAQBAJ&oi=fnd&pg=PP1&dq=data+preparation+techniques+machine+learning&ots=Cl3FsbaQxR&sig=Vi-juWPicuxR2FrZhFgANgs8Cs0#v=onepage&q=data preparation techniques machine learning&f=false

Bujold, E., Roberge, S., Lacasse, Y., Bureau, M., Audibert, F., Marcoux, S., Forest, J. C., & Giguère, Y. (2010). Prevention of preeclampsia and intrauterine growth restriction with aspirin started in early pregnancy: A meta-analysis. In *Obstetrics and Gynecology* (Vol. 116, Issue 2, pp. 402–414). https://doi.org/10.1097/AOG.0b013e3181e9322a

Burgen, M. (2018). *Google AI Could Help Predict When You Will Die*. Bloomberg. https://www.bloomberg.com/news/articles/2018-06-18/google-is-training-machines-to-predict-when-a-patient-will-die#xj4y7vzkg

Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo, J. L., Jones, D. W., Materson, B. J., Oparil, S., Wright, J. T., & Roccella, E. J. (2003). The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: The JNC 7 Report. In *Journal of the American Medical Association* (Vol.

289, Issue 19, pp. 2560–2572). American Medical Association. https://doi.org/10.1001/jama.289.19.2560

Clarke, M. (2022). *Critical shortage of doctors in SA*. https://www.da.org.za/2022/05/critical-shortage-of-doctors-in-sa-less-than-1-doctor-for-every-1-000-patients

Cunningham, P., Cord, M., & Delany, S. J. (2008). Supervised learning. *Cognitive Technologies*, 21–49. https://doi.org/10.1007/978-3-540-75171-7_2

Dantas, J. (2020). *The importance of k-fold cross-validation for model prediction in machine learning | by Jaime Dantas | Towards Data Science*. Towards Data Science. https://towardsdatascience.com/the-importance-of-k-fold-cross-validation-for-model-prediction-in-machine-learning-4709d3fed2ef

De Bernis, L., Kinney, M. V., Stones, W., Ten Hoope-Bender, P., Vivio, D., Leisher, S. H., Bhutta, Z. A., Gülmezoglu, M., Mathai, M., Belizán, J. M., Franco, L., McDougall, L., Zeitlin, J., Malata, A., Dickson, K. E., & Lawn, J. E. (2016). Stillbirths: Ending preventable deaths by 2030. *The Lancet*, *387*(10019), 703–716. https://doi.org/10.1016/S0140-6736(15)00954-X

De Leeuw, E. D. (2005). *To Mix or Not to Mix Data Collection Modes in Surveys*. www.websm.org

Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising qualitative and quantitative evidence: A review of possible methods. *Journal of Health Services Research and Policy*, *10*(1), 45–53. https://doi.org/10.1258/1355819052801804

Donders, A. R. T., van der Heijden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, *59*(10), 1087–1091. https://doi.org/10.1016/J.JCLINEPI.2006.01.014

Ebonwu, J., Mumbauer, A., Uys, M., Wainberg, M. L., & Medina-Marino, A. (2018). Determinants of late antenatal care presentation in rural and peri-urban communities in South Africa: A cross-sectional study. *PLoS ONE*, *13*(3). https://doi.org/10.1371/journal.pone.0191903

Finlayson, K., & Downe, S. (2013). Why Do Women Not Use Antenatal Services in Low- and Middle-Income Countries? A Meta-Synthesis of Qualitative Studies. *PLoS Medicine*, *10*(1). https://doi.org/10.1371/JOURNAL.PMED.1001373

Frizzell, J. D., Liang, L., Schulte, P. J., Yancy, C. W., Heidenreich, P. A., Hernandez, A. F., Bhatt, D. L., Fonarow, G. C., & Laskey, W. K. (2017). Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: Comparison of machine learning and other statistical approaches. *JAMA Cardiology*, *2*(2), 204–209. https://doi.org/10.1001/jamacardio.2016.3956

Gardosi, J., Madurasinghe, V., Williams, M., Malik, A., & Francis, A. (2013). Maternal and fetal risk factors for stillbirth: Population based study. *BMJ (Online)*, *346*(7893). https://doi.org/10.1136/bmj.f108

Garovic, V. D., & Hayman, S. R. (2007). Hypertension in pregnancy: an emerging risk factor for cardiovascular disease. *Nature Clinical Practice. Nephrology*, *3*(11), 613–622. https://doi.org/10.1038/ncpneph0623

Glaros, C., & I. Fotiadis, D. (2005). *Wearable Devices in Healthcare* (pp. 237–264). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11311966_8

Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. A. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *Journal of the American Medical Informatics Association*, *24*(1), 198–208. https://doi.org/10.1093/jamia/ocw042

Haas, D. M., Parker, C. B., Wing, D. A., Parry, S., Grobman, W. A., Mercer, B. M., Simhan, H. N., Hoffman, M. K., Silver, R. M., Wadhwa, P., Iams, J. D., Koch, M. A., Caritis, S. N., Wapner, R. J., Esplin, M. S., Elovitz, M. A., Foroud, T., Peaceman, A. M., Saade, G. R., … Reddy, U. M. (2015). A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). *American Journal of Obstetrics and Gynecology*, *212*(4), 539.e1-539.e24. https://doi.org/10.1016/J.AJOG.2015.01.019

Haddad, B., Deis, S., Goffinet, F., Paniel, B. J., Cabrol, D., & Sibaï, B. M. (2004). Maternal and perinatal outcomes during expectant management of 239 severe preeclamptic women between 24 and 33 weeks' gestation. *American Journal of Obstetrics and Gynecology*, *190*(6), 1590–1595. https://doi.org/10.1016/J.AJOG.2004.03.050

Harper, L., Kalfa, N., Beckers, G. M. A., Kaefer, M., Nieuwhof-Leppink, A. J., Fossum, M., Herbst, K. W., & Bagli, D. (2020). The impact of COVID-19 on research. *Journal of Pediatric Urology*, *16*(5), 715–716. https://doi.org/10.1016/J.JPUROL.2020.07.002

Hauger, M. S., Gibbons, L., Vik, T., & Belizán, J. M. (2008). Prepregnancy weight status and the risk of adverse pregnancy outcome. *Acta Obstetricia et Gynecologica Scandinavica*, *87*(9), 953–959. https://doi.org/10.1080/00016340802303349

Hawkins, D. M. (2004). The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences*, *44*(1), 1–12. https://doi.org/10.1021/CI0342472

Hevner, A., & Chatterjee, S. (2010). Design Science Research in Information Systems. *Design Research in Information Systems*, *22*, 9–22. https://doi.org/10.1007/978-1-4419-5653-8

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly: Management Information Systems*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports 2022 12:1*, *12*(1), 1–9. https://doi.org/10.1038/s41598-022-09954-8

Hinton, G. (2018). Deep learning-a technology with the potential to transform health care. In *JAMA - Journal of the American Medical Association* (Vol. 320, Issue 11, pp. 1101–1102). American Medical Association. https://doi.org/10.1001/jama.2018.11100

Hofmeyr, G. J., & Mentrop, L. (2015). Time for 'basic antenatal care plus' in South Africa?'. In *South African Medical Journal* (Vol. 105, Issue 11, pp. 902–903). South African Medical Association. https://doi.org/10.7196/SAMJ.2015.v105i11.10186

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(03), 90–95. https://doi.org/10.1109/MCSE.2007.55

Hutcheon, J. A., Platt, R. W., Abrams, B., Himes, K. P., Simhan, H. N., & Bodnar, L. M. (2013). A weight-gain-for-gestational-age z score chart for the assessment of maternal weight gain in pregnancy. *The American Journal of Clinical Nutrition*, *97*(5), 1062–1067. https://doi.org/10.3945/AJCN.112.051706

Hypertension in pregnancy. (2013). *Obstetrics and Gynecology*, *122*(5), 1122–1131. https://doi.org/10.1097/01.AOG.0000437382.03963.88

"Iterative Learning-InsideAIML." (n.d.). Retrieved June 15, 2021, from https://insideaiml.com/blog/Iterative-Learning-292

Jamshidian, M., & Mata, M. (2007). Advances in Analysis of Mean and Covariance Structure when Data are Incomplete. *Handbook of Latent Variable and Related Models*, 21–44. https://doi.org/10.1016/B978-044452044-9/50005-7

Jhee, J. H., Lee, S., Park, Y., Lee, S. E., Kim, Y. A., Kang, S. W., Kwon, J. Y., & Park, J. T. (2019). Prediction model development of late-onset preeclampsia using machine learning-based methods. *PLoS ONE*, *14*(8), e0221202. https://doi.org/10.1371/journal.pone.0221202

Joblib: running Python functions as pipeline jobs. (2008). https://joblib.readthedocs.io/en/latest/

Johannesson, P., & Perjons, E. (2014). An introduction to design science. *An Introduction to Design Science*, *9783319106328*, 1–197. https://doi.org/10.1007/978-3-319-10632-8

Jones, D., & Gregor, S. (2007). The Anatomy of a Design Theory. *Journal of the Association for Information Systems*, *8*(5), 19. https://doi.org/10.17705/1jais.00129

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. In *Science* (Vol. 349, Issue 6245, pp. 255–260). American Association for the Advancement of Science. https://doi.org/10.1126/science.aaa8415

Khan, K. S., Wojdyla, D., Say, L., Gülmezoglu, A. M., & Van Look, P. F. (2006). WHO analysis of causes of maternal death: a systematic review. *Lancet (London, England)*, *367*(9516), 1066–1074. https://doi.org/10.1016/S0140-6736(06)68397-9

Lamminpää, R., Vehviläinen-Julkunen, K., Gissler, M., & Heinonen, S. (2012). Preeclampsia complicated by advanced maternal age: a registry-based study on primiparous women in Finland 1997-2008. *BMC Pregnancy and Childbirth*, *12*(1), 1–5. https://doi.org/10.1186/1471-2393-12-47/TABLES/2

Lantz, B. (2013). *Machine Learning with R* (1st ed., Vol. 1). Packt Publishing.

Lantz, B. (2019). *Machine Learning with R: Expert techniques for predictive modeling* (Third). Packt Publishing. https://books.google.co.za/books?hl=en&lr=&id=iNuSDwAAQBAJ&oi=fnd&pg=PP1&dq=how+to+perform+machine+learning+models&ots=O86Xxd5CQ-&sig=DwHnMIFUdPprPXomzpqez4uy0mA#v=onepage&q=how to perform machine learning models&f=false

Lawn, J. E., Blencowe, H., Waiswa, P., Amouzou, A., Mathers, C., Hogan, D., Flenady, V., Frøen, J. F., Qureshi, Z. U., Calderwood, C., Shiekh, S., Jassir, F. B., You, D., McClure, E. M., Mathai, M., Cousens, S., Kinney, M. V., De Bernis, L., Heazell, A., … Draper, E. S. (2016). Stillbirths: Rates, risk factors, and acceleration towards 2030. *The Lancet*, *387*(10018), 587–603. https://doi.org/10.1016/S0140-6736(15)00837-5

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In *Nature* (Vol. 521, Issue 7553, pp. 436–444). Nature Publishing Group. https://doi.org/10.1038/nature14539

Lindheimer, M. D., Taler, S. J., & Cunningham, F. G. (2008). Hypertension in pregnancy. *Journal of the American Society of Hypertension*, *2*(6), 484–494. https://doi.org/10.1016/J.JASH.2008.10.001

Madazli, R., Yuksel, M. A., Imamoglu, M., Tuten, A., Oncul, M., Aydin, B., & Demirayak, G. (2014). Comparison of clinical and perinatal outcomes in early- and late-onset preeclampsia. *Archives of Gynecology and Obstetrics*, *290*(1), 53–57. https://doi.org/10.1007/S00404-014-3176-X

Madhi, S. A., Briner, C., Maswime, S., Mose, S., Mlandu, P., Chawana, R., Wadula, J., Adam, Y., Izu, A., & Cutland, C. L. (2019). Causes of stillbirths among women from South Africa: a prospective, observational study. *The Lancet Global Health*, *7*(4), e503–e512. https://doi.org/10.1016/S2214-109X(18)30541-2

Maphumulo, W. T., & Bhengu, B. R. (2019). Challenges of quality improvement in the healthcare of South Africa post-apartheid: A critical review. In *Curationis* (Vol. 42, Issue 1). NLM (Medline). https://doi.org/10.4102/curationis.v42i1.1901

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2

Maternal mortality. (n.d.). Retrieved June 15, 2021, from https://www.who.int/news-room/fact-sheets/detail/maternal-mortality

Mobbs, R. J., Ho, D., Choy, W. J., Betteridge, C., & Lin, H. (2020). COVID-19 is shifting the adoption of wearable monitoring and telemedicine (WearTel) in the delivery of healthcare: opinion piece. *Annals of Translational Medicine*, *8*(20), 1285–1285. https://doi.org/10.21037/atm-20-3678

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning, second edition - Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar - Google Books*. The MIT Press. https://books.google.co.za/books?hl=en&lr=&id=dWB9DwAAQBAJ&oi=fnd&pg=PR5&dq=machine+learning&ots=AyoQTSr2r0&sig=uA9coMnaohutimIV6uq0tSFo7ww#v=onepage&q&f=false

Morgenthaler, S. (2009). Exploratory data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(1), 33–44. https://doi.org/10.1002/WICS.2

Müller, A. C., & Guido, S. (2017). *Introduction to Machine Learning with Python: A Guide for Data Scientists* (First). O'Reilly Media. https://books.google.co.za/books?hl=en&lr=&id=1-4lDQAAQBAJ&oi=fnd&pg=PP1&dq=machine+learning+guide&ots=28lUKKJMTW&sig=P3FZdGLp2hTGqEY-QxoNktJ3Q-g#v=onepage&q=machine learning guide&f=false

Myer, L., & Harrison, A. (2003). Why do womn seek antenatal care late? Perspectives from rural South Africa. *Journal of Midwifery and Women's Health*, *48*(4), 268–272. https://doi.org/10.1016/S1526-9523(02)00421-X

Myers, J. E., Myatt, L., Roberts, J. M., & Redman, C. W. G. (2019). COLLECT, a collaborative database for pregnancy and placental research studies worldwide. In *BJOG: An International Journal of Obstetrics and Gynaecology* (Vol. 126, Issue 1, pp. 8–10). Blackwell Publishing Ltd. https://doi.org/10.1111/1471-0528.15393

Myers, M. D., & Venable, J. R. (2014). A set of ethical principles for design science research in information systems. *Information & Management*. https://doi.org/10.1016/j.im.2014.01.002

Nathan, H. L., Seed, P. T., Hezelgrave, N. L., De Greeff, A., Lawley, E., Conti-Ramsden, F., Anthony, J., Steyn, W., Hall, D. R., Chappell, L. C., & Shennan, A. H. (2018). Maternal and

perinatal adverse outcomes in women with pre-eclampsia cared for at facility-level in South Africa: A prospective cohort study. *Journal of Global Health*, *8*(2). https://doi.org/10.7189/jogh.08.020401

NICE. (2019). *Hypertension in pregnancy: diagnosis and management*. www.nice.org.uk/guidance/ng133

North, R. A., McCowan, L. M. E., Dekker, G. A., Poston, L., Chan, E. H. Y., Stewart, A. W., Black, M. A., Taylor, R. S., Walker, J. J., Baker, P. N., & Kenny, L. C. (2011). Clinical risk prediction for pre-eclampsia in nulliparous women: development of model in international prospective cohort. *BMJ*, *342*(7803). https://doi.org/10.1136/BMJ.D1875

Oliphant, T. E. (2006). *Guide to NumPy*. http://www.trelgol.com

Pandas - A Python Data Analysis Library. (2008). https://pandas.pydata.org/about/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhfer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., & Duchesnay, M. P. E. (2011). Scikit-learn: Machine Learning in Python. In *Journal of Machine Learning Research* (Vol. 12). http://scikit-learn.sourceforge.net.

Peffers, K., Tuunanen, T., & Gengler, C. (2006). The design science research process: a model for producing and presenting information systems research. *Journal of Management Information Systems*, *February*, 83–106.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pries-Heje, J., Baskerville, R., & Venable, J. R. (2008). *Strategies for Design Science Research Evaluation*. *87*. http://aisel.aisnet.org/ecis2008/87

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, *380*(14), 1347–1358. https://doi.org/10.1056/NEJMRA1814259

Reinebrant, H., Leisher, S., Coory, M., Henry, S., Wojcieszek, A., Gardener, G., Lourie, R., Ellwood, D., Teoh, Z., Allanson, E., Blencowe, H., Draper, E., Erwich, J., Frøen, J., Gardosi, J., Gold, K., Gordijn, S., Gordon, A., Heazell, A., … Flenady, V. (2018). Making stillbirths visible: a systematic review of globally reported causes of stillbirth. *BJOG: An International Journal of Obstetrics & Gynaecology*, *125*(2), 212–224. https://doi.org/10.1111/1471-0528.14971

Rezk, M., Gamal, A., & Emara, M. (2015). Maternal and fetal outcome in de novo preeclampsia in comparison to superimposed preeclampsia: a two-year observational study. *Hypertension in Pregnancy*, *34*(2), 137–144. https://doi.org/10.3109/10641955.2014.982329

Roberts, J. M., & Gammill, H. S. (2005). Preeclampsia: Recent insights. In *Hypertension* (Vol. 46, Issue 6, pp. 1243–1249). Lippincott Williams & Wilkins. https://doi.org/10.1161/01.HYP.0000188408.49896.c5

Rosser, M. L., & Katz, N. T. (2013). Preeclampsia: An Obstetrician's Perspective. In *Advances in Chronic Kidney Disease* (Vol. 20, Issue 3, pp. 287–296). W.B. Saunders. https://doi.org/10.1053/j.ackd.2013.02.005

Rusanov, A., Weiskopf, N. G., Wang, S., & Weng, C. (2014). Hidden in plain sight: Bias towards sick patients when sampling patients with sufficient electronic health record data for research. *BMC Medical Informatics and Decision Making*, *14*(1), 1–9. https://doi.org/10.1186/1472-6947-14-51

Ruxwana, N. L., Herselman, M. E., & Conradie, D. P. (2010). ICT applications as e-health solutions in rural healthcare in the Eastern Cape Province of South Africa. *The HIM Journal*, *39*(1), 17–26. https://doi.org/10.1177/183335831003900104

Savaj, S., & Vaziri, N. (2012). *An overview of recent advances in pathogenesis and diagnosis of preeclampsia.* https://escholarship.org/uc/item/3863p1k7

Seneviratne, S., Hu, Y., Nguyen, T., Lan, G., Khalifa, S., Thilakarathna, K., Hassan, M., & Seneviratne, A. (2017). A Survey of Wearable Devices and Challenges. In *IEEE Communications Surveys and Tutorials* (Vol. 19, Issue 4, pp. 2573–2620). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/COMST.2017.2731979

Shen, M., Smith, G. N., Rodger, M., White, R. R., Walker, M. C., & Wen, S. W. (2017). Comparison of risk factors and outcomes of gestational hypertension and pre-eclampsia. *PLoS ONE*, *12*(4). https://doi.org/10.1371/JOURNAL.PONE.0175914

Statistics South Africa. (2020). *Maternal Health Indicators*. http://www.statssa.gov.za/?p=13100

Swamynathan, M. (2017). Mastering Machine Learning with Python in Six Steps. In *Mastering Machine Learning with Python in Six Steps*. Apress. https://doi.org/10.1007/978-1-4842-2866-1

Sylvia Patience Ngxongo, T. (2019). Basic Antenatal Care Approach to Antenatal Care Service Provision. In *Selected Topics in Midwifery Care*. IntechOpen. https://doi.org/10.5772/intechopen.79361

T Akinsola, J. E., Jet, A., & O, H. J. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology*, *48*. https://doi.org/10.14445/22312803/IJCTT-V48P126

Tranquilli, A. L., Dekker, G., Magee, L., Roberts, J., Sibai, B. M., Steyn, W., Zeeman, G. G., & Brown, M. A. (2014). The classification, diagnosis and management of the hypertensive disorders of pregnancy: A revised statement from the ISSHP. In *Pregnancy Hypertension* (Vol. 4, Issue 2, pp. 97–104). Elsevier B.V. https://doi.org/10.1016/j.preghy.2014.02.001

Trochim, W. M. K. (2006). *Research Design - Research Methods Knowledge Base*. https://conjointly.com/kb/research-design/

Trudell, A. S., Tuuli, M. G., Colditz, G. A., Macones, G. A., & Odibo, A. O. (2017). A stillbirth calculator: Development & internal validation of a clinical prediction model to quantify stillbirth risk. *PLoS ONE*, *12*(3), 1–13. https://doi.org/10.1371/journal.pone.0173461

Tucker, K. L., Bankhead, C., Hodgkinson, J., Roberts, N., Stevens, R., Heneghan, C., Rey, É., Lo, C., Chandiramani, M., Taylor, R. S., North, R. A., Khalil, A., Marko, K., Waugh, J., Brown, M., Crawford, C., Taylor, K. S., Mackillop, L., & McManus, R. J. (2018). How do home and clinic blood pressure readings compare in pregnancy? A systematic review and individual patient data meta-analysis. In *Hypertension* (Vol. 72, Issue 3, pp. 686–694). Lippincott Williams and Wilkins. https://doi.org/10.1161/HYPERTENSIONAHA.118.10917

Tucker, K. L., Taylor, K. S., Crawford, C., Hodgkinson, J. A., Bankhead, C., Carver, T., Ewers, E., Glogowska, M., Greenfield, S. M., Ingram, L., Hinton, L., Khan, K. S., Locock, L., Mackillop, L., McCourt, C., Pirie, A. M., Stevens, R., & McManus, R. J. (2017). Blood pressure self-monitoring in pregnancy: Examining feasibility in a prospective cohort study.

BMC Pregnancy and Childbirth, *17*(1), 442. https://doi.org/10.1186/s12884-017-1605-0

UNICEF DATA. (2021). *Antenatal care*. https://data.unicef.org/topic/maternal-health/antenatal-care/

Vaishnavi, V., & Kuechler, B. (2004). *Design Research in Information Systems*. http://www.desrist.org/design-research-in-information-systems/

van de Leur, R. R., Boonstra, M. J., Bagheri, A., Roudijk, R. W., Sammani, A., Taha, K., Doevendans, P. A. F. M., van der Harst, P., van Dam, P. M., Hassink, R. J., van Es, R., & Asselbergs, F. W. (2020). Big data and artificial intelligence: Opportunities and threats in electrophysiology. *Arrhythmia and Electrophysiology Review*, *9*(3), 146–154. https://doi.org/10.15420/AER.2020.26

van den Heuvel, J. F. M., Kariman, S. S., van Solinge, W. W., Franx, A., Lely, A. T., & Bekker, M. N. (2019). SAFE@HOME – Feasibility study of a telemonitoring platform combining blood pressure and preeclampsia symptoms in pregnancy care. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, *240*, 226–231. https://doi.org/10.1016/j.ejogrb.2019.07.012

van der Merwe, A., Gerber, A., & Smuts, H. (2017). Mapping a design science research cycle to the postgraduate research report. *Communications in Computer and Information Science*, *730*, 293–308. https://doi.org/10.1007/978-3-319-69670-6_21

Venable, J., Pries-Heje, J., & Baskerville, R. (2012). A Comprehensive Framework for Evaluation in Design Science Research. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7286 LNCS*, 423–438. https://doi.org/10.1007/978-3-642-29863-9_31

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. *Information Systems Research*, *3*(1), 36–59. https://doi.org/10.1287/ISRE.3.1.36

Weir, C. B., & Jan, A. (2021). BMI Classification Percentile And Cut Off Points. *StatPearls*. https://www.ncbi.nlm.nih.gov/books/NBK541070/

Wildschut, A. (2010). Doctors in the public service: too few for too many. *HSRC Review*, *8*, 12–15. http://www.hsrc.ac.za/en/review/november-/public-service-doctors

Williams, C. (2007). Research Methods. *Journal of Business & Economics Research (JBER)*, *5*(3), 65. https://doi.org/10.19030/JBER.V5I3.2532

World Health Organisation. (2016). *WHO: Stillbirths*. World Health Organization. https://www.who.int/maternal_child_adolescent/epidemiology/stillbirth/en/

Wu, Y., Moylan, E., Inman, H., & Graf, C. (2019). Paving the Way to Open Data. *Data Intelligence*, *1*(4), 368–380. https://doi.org/10.1162/DINT_A_00021

Zhang, S., Zhang, C., & Yang, Q. (2016). *Data preparation for data mining*. https://doi.org/10.1080/713827180

Nathan Smith is a second-year MSc student in the Department of Computing Sciences at the Nelson Mandela University. His research interests are machine learning, telemedicine, and smart spaces.

APPENDIX D: GitHub repository for all code for the development of the models

The code used in the development of the models can be found at the following link: https://github.com/Celibral/Nathan-Smith-Master-s-Project

APPENDIX E: Sample of Iteration 1's dataset

	Starting_	Height	Parity	Multiple	HT_Hist	Previous	Previou	Diabetic	Diabete	Age	HT_Fhis	Week_1	Mass_1	Week_2	Mass_2	Week_3	Mass_3	Week_4	Mass_4	Develop	Starting	Starting	Starting	SBP_1	DBP_1	SBP_2	DBP_2	SBP_3	DBP_3	SBP_4	DBP_4
0	50.9	156	0	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	21	FALSE	22	54	26	58	31	62	0	0	FALSE	16	120	70	110	70	120	70	110	70	0	0
1	79.6	0	0	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	30	FALSE	13	82	17	85	23	87.3	27	90.2	FALSE	8	134	81	127	76	130	80	112	71	123	77
2	0	0	0	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	28	FALSE	17	75	24	80	26	80	32	89.2	TRUE	9	127	82	116	78	114	77	112	67	147	108
3	70	173	4	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	26	FALSE	25	77	33	78.8	36	78.8	0	0	FALSE	13	129	75	110	70	124	69	135	68	122	77
4	50	0	0	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	23	FALSE	18	52	22	57	26	60	29	63	FALSE	10	0	0	132	75	113	66	110	62	109	61
5	76.6	165	0	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	32	FALSE	18	77.9	22	80	25	83	30	85.2	FALSE	9	125	85	129	82	117	78	118	774	109	75
6	0	175	1	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	33	FALSE	16	61.4	23	66.5	28	67.9	31	69.8	FALSE	15	124	61	124	61	115	65	122	63	107	61
7	92	173	0	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	30	FALSE	13	95.3	16	96	23	100	33	106.2	FALSE	7	110	64	120	70	109	68	115	60	102	55
8	68.7	170	0	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	38	FALSE	16	71	24	74	28	75	36	78.5	FALSE	12	0	0	108	66	100	64	115	70	119	78

APPENDIX F: Correlation Matrix for all features in Iteration One's Dataset