Adverse Pregnancy Outcome Prediction Using Deep Learning

Fozia Abako Dengo

A Thesis Submitted to the Department of Computer Science and Engineering

School of Electrical Engineering and Computing

Presented in partial fulfillment of the requirement for Degree of Master's in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

September 2023

Adama, Ethiopia

Adverse Pregnancy Outcome Prediction Using Deep Learning

Fozia Abako Dengo

Advisor: Dr. Mesfin Abebe

A Thesis Submitted to the Department of Computer Science and Engineering
School of Electrical Engineering and Computing

Presented in Partial Fulfillment of the Requirement for Degree of Master's in
Computer Science and Engineering

Office of Graduate Studies
Adama Science and Technology University

September 2023
Adama, Ethiopia

DECLARATION

I hereby declare that this Master Thesis entitled "**Adverse Pregnancy Outcome Prediction Using Deep Learning**" is my original work. That is, it has not been submitted for the award of any academic degree, diploma, or certificate in any other university. All sources of materials that are used for this thesis have been duly acknowledged through citation

Fozia Abako Dengo	______________________	_______________
Name of the student	Signature	Date

RECOMMENDATION

I, the advisor of this thesis, hereby certify that I have read the revised version of the thesis entitled "**Adverse Pregnancy Outcome Prediction Using Deep Learning**" prepared under my guidance by **Fozia Abako Dengo** submitted in partial fulfillment of the requirements for the degree of Masters of Science in Computer Science and Engineering. Therefore, I recommend the submission of a revised version of the thesis to the department following the applicable procedures.

Dr. Mesfin Abebe_	________________________	________________
Major Advisor	Signature	Date

APPROVAL SHEET

I the advisor of the thesis entitled "**Adverse Pregnancy Outcome Prediction Using Deep Learning**" and developed by "**Fozia Abako Dengo**" hereby certify that the recommendation and suggestion made by the board of examiners are appropriately incorporated into the final version of the thesis.

Dr. Mesfin Abebe	______________________	__________________________
Major Advisor/Supervisor	Signature	Date

We, the undersigned, members of the Board of Examiners of the thesis by Fozia Abako Dengo have read and evaluated the thesis entitled "**Adverse Pregnancy Outcome Prediction Using Deep Learning**" and examined the candidate during the open defense. This is, therefore, to certify that the thesis is accepted for partial fulfillment of the requirement of the degree of Masters of Science in Computer Science and Engineering.

____________________	____________________	__________________
Chairperson	Signature	Date
____________________	____________________	__________________
Reviewer 1	Signature	Date
____________________	____________________	__________________
Reviewer 2	Signature	Date

Final approval and acceptance of the thesis are contingent upon submission of its final copy to the Office of Postgraduate Studies (OPGS) through the Department Graduate Council (DGC) and School Graduate Committee (SGC).

____________________	____________________	____________________
Department Head	Signature	Date
____________________	____________________	____________________
School Dean	Signature	Date
____________________	____________________	____________________
Office of Postgraduate Studies, Dean	Signature	Date

ACKNOWLEDGMENT

Beyond my academic endeavors, I am eternally indebted to your boundless love and support. With a heart overflowing with gratitude and reverence, I offer the acknowledgment to ALLAH for the immeasurable love, unconditional favor, and support ALLAH has provided me throughout the completion of the thesis and my life.

I would like to thank my beloved Mother Alemitu Belda for her selfless love, unwavering support, countless prayers, and encouragement. Her unconditional support and belief in my abilities are the only things that kept me inspired and craving to do more. This accomplishment wouldn't have been possible without her endless prayers and faith in me.

I am profoundly grateful to my esteemed advisor Dr. Mesfin Abebe whose mentorship and expertise have been pivotal throughout the thesis journey. I am fortunate to have had the opportunity to work under your supervision.

Additionally, I would like to express my gratitude and heartfelt appreciation to my dearest friend Dureti Moti. Your unwavering support, our brain-storming sessions, and words of encouragement were the constant source of motivation. I would like to express my gratitude to my friend Natnael Girma who generously dedicated his time and shared his insights for this study. Your presence and friendship throughout the study have made the completion of the study valuable and memorable

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A ALRR : ADD-Left Remove-Right selection ANN : Artificial Neural Network AUC : Area Under Curve APO: Adverse Pregnancy Outcome APOP: Adverse pregnancy outcome Prediction	K K-NN : K-neighbor Nearest
B BiLSTM: Bidirectional Long Term Short Memory	L LSTM : Long Short-term Memory LASSO: Least Absolute Shrinkage and Selection Operator LR-S: Logistic regression with Stepwise Selection LR: Logistic Regression
C CNN: Conventional Neural Network CDSS: Clinical decision support system CB: Cast Boost	M MLM : Machine Learning Method MLP : Multi-layer Perceptron ML : Machine Learning MLA: Machine Learning Approach MLNN: Multi-Layer Neural Network
D DT: Decision Tree DL: Deep Learning DLA: Deep Learning Approach	N NN: Neural Network NFPC: National Free Pre-Pregnancy Check-ups
E EHG: Electro hysterogram EGB: Extreme Gradient Boosting	R RF: Random Forest RNN: Recurrent Neural Networks
F FNN: Feed Forward Neural Network	S SVM : Support Vector Machine SMOTE: Synthetic Minority Over-sampling Technique SL: Super Learner SLE: Systemic Lupus Erythematosus

ABSTRACT

Pregnancy outcome, also known as birth outcome, refers to the long-term effects of fertilization on the newborn child, extending from the beginning of fetal viability (about 28 weeks gestation) to the first weeks following birth. Adverse pregnancy outcome is an encompassing term that comprises health concerns that affect either the expectant mother, the baby, or both, and throughout the whole pregnancy, childbirth, and the postpartum period. Nearly, 94% of these complications are treatable and preventable. Ethiopia ranks among the nations that has the highest number of maternal fatalities. pregnant women from developing countries are thirty-six times more prone to experience complications associated with pregnancy than those of developed countries. Early prediction of these pregnancy complications has a strong correlation with women's survival. This study was carried out in accordance with experimental and design science approaches. The objective of this study, was to develop a deep-learning-based model to predict adverse pregnancy outcomes. The model construction process involved conducting five experiments, each of which utilized a dataset comprising a total of 10,220 instances and featured 15 variables, one of which is a target variable. We have developed a dataset for model development gathered from EDHS. The data collected from EDHS were prepared and preprocessed by using various preprocessing techniques to get quality data that is suitable for the proposed deep-learning approaches to predict adverse pregnancy outcomes. To construct the predictive model, we trained distinct models, by using deep learning algorithms such as LSTM, FNN, BiLSTM, FNN_LSTM, and FNN_BiLSTM for comparison of their performance of prediction of adverse pregnancy outcomes. We have examined the effect of algorithm hybridizations by using FNN with LSTM and BiLSTM. The performance of the model is assessed under stratified 10-fold cross-validation. Based on the experiment performed the FNN_BiLSTM outperformed other models for the prediction of adverse pregnancy outcomes. It achieved 98.94% of accuracy with 99% of F1-Score. The proposed predictive model classifies adverse pregnancy outcomes based on a multi-class prediction approach. Finally, we recommend future researchers to focus on enhancing the predictive model by utilizing a substantial dataset for more extensive investigation.

Key word: Adverse pregnancy outcome, Deep Learning Models, Ethiopian demographic health survey, pregnancy

CHAPTER ONE

1. INTRODUCTION

1.1 Background of the Study

Pregnancy is a complex vital period in a woman's life that can have profound effects on both her physical and mental well-being. Adapting to the significant physiological changes that take place during pregnancy can pose a challenge. Moreover, prioritizing both her personal well-being as well as her unborn child is a fundamental aspect of her psychological wellness, encompassing her emotional conduct (Andreea M. Oprescu, 2020).

According to the report from women's health office (Services, 2021)[1] Pregnancy is conventionally divided in to three trimesters each roughly three months long. In the first trimester of pregnancy (weeks 1-12), any harm to the fetus carries a high risk of causing miscarriage or severe, long-term disabilities. During the second trimester of pregnancy (weeks 13-28), the fetus experiences notable developments as its physical structures become more defined and start functioning to a certain degree. In the third trimester of pregnancy (weeks 29-40), the fetus experiences significant growth and prepares for birth (WHO, 2019) (Fowler, Mahdy, Jack, & W., 2023).

Pregnancy outcome, also known as birth outcome, refers to the long-term effects of fertilization on the newborn child, extending from the beginning of fetal viability (about 28 weeks gestation) to the first weeks following birth. The results of each pregnancy can vary, encompassing various outcomes such as premature birth, stillbirth, spontaneous abortion, medically induced abortion, and early neonatal death (Yeshialem E., 2019).

Adverse pregnancy outcome is an encompassing term that comprises health concerns that affect either the expectant mother, the baby, or both, and throughout the whole pregnancy, childbirth, and the postpartum period. Common pregnancy complications encompass a range of issues, including bleeding during pregnancy, hyperemesis gravidarum, postpartum hemorrhage, stillbirth, low birth weight, early rupture of the amniotic sac, obstructed labor, hypertensive disorders during pregnancy, premature birth, uterine rupture, and puerperal sepsis.(Mesfin

[1] www.womenshealth.gov/pregnancy/youre-pregnant-now-what/stages-pregnancy

Tadese, 2022). Regardless of how frequently they occur, there are several ways of dealing with the problems that come up throughout pregnancy (System, 2021)[2].In order to reduce the likelihood of adverse pregnancy outcomes, women should seek medical care before and during their pregnancy.

In the developing world, approximately 300 million women experience either short-term or long-term health issues as a result of pregnancy and childbirth. The majority of maternal deaths occur in these developing regions. In developing countries, maternal complications during pregnancy and childbirth continue to be significant contributors to both mortality and disability among women of reproductive age (WHO, 2019)**.**

Despite the fact that the majority of pregnancies and deliveries are uncomplicated, every pregnancy has a certain amount of risk. Based on data from the (WHO, 2019)[3] report, approximately 15% of pregnant women will experience a severe complication during pregnancy, necessitating specialized medical attention, with some cases demanding significant obstetric intervention for the mother's survival. Around 800 women die every day worldwide due to preventable complications associated with the inherent dangers of pregnancy. About 295,000 women die during and throughout pregnancy and childbirth in 2017 in a related report according to WHO. Nearly all of these fatalities (94%) occurred in areas with minimal resources, and a sizeable fraction of them may have been prevented (WHO U. U., 2016)[4] (Lale, 2015).

Mothers who delivered their babies at home faced a fourfold higher risk of experiencing adverse pregnancy outcomes compared to mothers who received care in a healthcare facility. Mothers who delivered newborns with a birth weight below 2.5 kilograms faced a 63% higher chance of experiencing pregnancy complications compared to women who delivered babies weighing between 2.5 kilograms and 3.99 kilograms (Tadese M, 2022). Prevention can be done by increasing awareness of pregnancy complication signs, which has a strong correlation with early prediction of pregnancy risks. Women who know the pregnancy complication signs are at 6.657 times more likely to understand about early prediction of pregnancy risks than those who do not understand about pregnancy risks (Mardiyanti I, 2019). Understanding the signs of pregnancy

[2] https://maternity.jacksonhealth.org/pregnancy-milestones/
[3] https://www.who.int/publications/i/item/WHO-MCA-17.02
[4] https://www.who.int/news-room/fact-sheets/detail/maternal-mortality

complications is significantly linked to receiving proper antenatal care. In essence, being aware of these signs plays a pivotal role in accurately predicting early pregnancy risks (Wassihun B, 2020) (Mardiyanti I, 2019). The frequency of maternal deaths has decreased (WHO, 2021) as a result of recent technological improvements. However, ensuring the safety of the mother and infant throughout pregnancy is still challenging. In such circumstances, the risks during pregnancy might be reduced by anticipating complications and taking preventative measures. The frequency of maternal deaths has decreased as a result of recent technological improvements. The use of predictive modeling has become crucial in preserving the lives of countless women and newborns. Deep learning has taken on a significant position within the realm of medical healthcare, with several application possibilities. Deep learning is being researched for use in medical imaging, electronic health records, genomics, and drug discovery. Deep learning provides a distinct edge in these areas by successfully exploiting biological data and increasing the standards of medical healthcare (Xang S, 2021). For health-related learning models, various kinds of data are used, including electronic medical records, medical images, biochemical measures, and biological markers (Ahmed, 2020).

This study aimed to classify adverse pregnancy outcomes using deep learning approaches. Therefore, maternal healthcare-related data were collected from the Ethiopian demographic health survey and classified three common adverse pregnancy outcomes such as low birth weight, preterm birth, and normal birth.

1.2 Motivation of the Study

Women can face life-threatening complications during pregnancy and childbirth, most of these complications are preventable or treatable. Birth preparedness and complications readiness aim to reduce delays in accessing skilled care by helping expectant mothers plan for childbirth and potential complications. Despite efforts, some complications during pregnancy are inevitable, leading to low birth weight or genetically malformed infants (Kamineni V, 2017).

According to (Stephen G, 2018), anemia during pregnancy is one of the risk factors for poor pregnancy outcomes such as low birth weight, preterm birth, stillbirth, intrauterine growth restriction, and impaired cognitive development. Risk factors for adverse maternal health outcomes during pregnancy include a variety of factors such as age, pre-existing medical conditions, lifestyle factors and socio-economic status (Crear-Perry, 2021) (Londero, 2019).

Based on the problems that are identified the researcher was motivated to work on adverse pregnancy outcome Prediction using deep learning approaches by utilizing and classifying maternal health care data that were collected from the EDHS dataset. Moreover, the study aims to propose a predictive model for adverse pregnancy outcomes using deep-learning approaches in order to overcome the problems stated above. In fact, predicting adverse pregnancy outcomes on their early stage will reduce the suffer of expectants families from physical, psychological and emotional crisis and it will reduce the maternal mortality rate.

1.3 Statement of the Problem

Every day, 800 women die from complications of pregnancy and childbirth. A pregnant woman residing in a developing country faces a significantly higher risk of experiencing pregnancy-related complications, with a 36-fold increase compared with a pregnant woman in a developed country (WHO, 2021).

There are various risk factors related to adverse pregnancy outcomes in Ethiopia. Among these factors, adverse pregnancy outcomes stem from inadequate prenatal care quality, and postpartum care, a lack of emphasis on prevention measurements, a failure to acknowledge the societal repercussions of adverse pregnancy outcomes, and a faulty data collection system (Al-Hindi, 2020). Many adverse pregnancy outcomes could have been prevented if women had received high-quality personalized care before conception, during pregnancy, and delivery (Hug L, 2020).

The majority of mothers die in their homes because of a lack of understanding and characterization of pregnancy complication signs (Abas AA, 2017). As a result, women from developing countries or those without access to high-quality healthcare may be more likely to experience pregnancy complications. Furthermore, pregnant women have consistently been left out of the majority of clinical trials, resulting in a lack of data regarding the effectiveness and safety of medications during pregnancy (van der Graaf, 2018).

Different studies (Tesfay, 2023) conducted on maternal healthcare fields have identified the factors that increase maternal complications. Of these factors, one is a lack of depth in exploring relationships among contributing attributes. For appropriate prevention of adverse pregnancy outcomes, data pertaining to the determinants of adverse pregnancy outcomes are important (Abadiga, 2020). According to (Hailu M, 2010) early prediction of these risk factors is critical for women’s survival. The occurrence of LBW in various studies conducted in Ethiopia ranged from 6 % to 29.1% (Alemu, 2016) (Zerfu, 2016). The overall prevalence of preterm birth conducted (Muchie KF, 2020) in Ethiopia was 10.48%. These indicate that Ethiopia is one of those countries that have the highest number of maternal fatalities.

The majority of the scholars who conducted their studies regarding the problems mentioned above had put their suggested solutions. One of these scholars (Hisham Allahem, 2022) mitigated the consequence of premature birth for pregnant women and fetuses using ML and DL. In this study, they have used a single feature to overcome the problem. The study failed to incorporate the most important risk factors. Important evaluation metrics were not included, as a matter of fact, the accuracy of the predictive model is less. Furthermore, the study only focused on preterm birth which makes it under a binary Prediction problem. Another study was conducted by (Yu Mu, 2018) to predict adverse pregnancy outcomes using deep learning approaches with the help of MLNN. This study has used a large feature set that makes feature engineering and selection challenging. Because managing and selecting relevant features from a large pool can be difficult. These affected the performance of the model negatively. In addition, optimization techniques were not incorporated to the fullest into it. (Lakshmi.B.N, 2015) have also conducted their study aiming to monitor women during pregnancy to protect them from possible pregnancy complications. In this work, a small amount of dataset was used and additional parameters and methods were not incorporated.

Therefore, to prioritize the precedence of adverse pregnancy outcomes among Ethiopian women and promote the improvement of an effective healthcare system, it is essential to emphasize the need for a predictive model that concentrates on developing secure and non-intrusive approaches. However, in the framework of our country, no one has developed a predictive model for adverse pregnancy outcomes specifically using the Ethiopian demographic and health survey dataset targeting Ethiopian women. This study proposed a deep-learning-based model to solve the above-mentioned problem in maternal healthcare by predicting adverse pregnancy outcomes using the EDHS dataset. Furthermore, the study allows and encourages healthcare practitioners to take proactive measures and offer tailored assistance.

1.4 Research Questions

This study will address the following research questions:

1. What are the most determining attributes that influence the occurrence of adverse pregnancy outcomes?
2. Which deep learning algorithm is best suited for developing a model for classifying adverse pregnancy outcomes?
3. What is the hyperparameter that has the most significant impact on the performance of the model during the experiment?

1.5 Objective of the Study

1.5.1 General Objective

The general objective of this research is to develop deep-learning based model that predicts adverse pregnancy outcomes.

1.5.2 Specific Objective

To achieve the general objective of the study, the following specific tasks are to be address

- Reviewing formerly directed research and literature on adverse pregnancy outcomes.
- Gathering maternal health-related data and preparing the dataset for the model development.
- Identifying the most relevant features and extraction techniques for the model development.
- Identifying the most suited deep learning algorithm for the model development
- Building a prediction model for adverse pregnancy outcomes.
- Evaluating the performance of the prediction model.

1.6 Significance of the Study

The prediction of adverse pregnancy outcomes using deep learning approaches can benefit various stakeholders in the health care system. First and foremost, expectant mothers and their families are among the primary beneficiaries. Precise and accurate prediction of adverse pregnancy outcomes can lead to early interventions and personalized care plans, potentially reducing the risks associated with complications. This has the potential to enhance maternal and neonatal health results and alleviate anxiety among expectant parents. Healthcare providers who specialize in maternal and fetal medicine can benefit from improved Prediction models. These can assist them in making more informed decisions regarding patient care, including identifying high-risk pregnancies, optimizing monitoring strategies, and planning appropriate interventions. Public health officials also benefit from these studies. Public health bureaus can use data from predictive models to identify trends and patterns in adverse pregnancy outcomes at a population level. This information can inform public health policies and interventions intended at reducing maternal and neonatal mortality and morbidity. Researchers and academia in the field of maternal-fetal medicine and artificial intelligence can use predictive models to advance the understanding of the risk factors and develop more effective prevention and treatment strategies. Healthcare systems can benefit from reduced healthcare costs associated with pregnancy complications. Early identification of high-risk pregnancies can lead to better resource allocation and reduce hospitalization and emergency care costs. Furthermore, society benefits from improved maternal and neonatal outcomes. Reducing adverse pregnancy outcomes can lead to healthier mothers, healthier babies, and improved quality of life. The outcome of this study can serve as input to different Prediction tasks about pregnancy and obstetric signs.

1.7 Scope and Limitation of the Study

1.7.1 Scope of the Study

The scope of this study focused on adverse pregnancy outcome prediction using deep learning models. In order to reach this goal, the data was collected from EDHS. The study focused on developing a deep learning model that predicts adverse pregnancy outcomes. To provide a predictive model, various number of deep learning models were proposed with their respective experimentations. The study also assessed the effectiveness of the proposed model by using performance evaluation metrics. This study also presented a hybrid combination of deep learning models to improve the performance of the model.

1.7.2 Limitation of the Study

The presented study focuses only in predicting adverse pregnancy outcome. Due to lack of data quality and imbalance constraints this study will classify only the presence of low birth weight, preterm birth, and normal birth. The other limitation when working with maternal health related dataset, in our country, no prior research has utilized this dataset to make comparisons regarding the contributing factors and effectiveness of the predictive model. The proposed study doesn’t diagnose or treat a disease. The proposed model only classifies adverse pregnancy outcomes of pregnant women about their respective feature.

1.8 Organization of the Thesis

The overall summary of the chapters in our thesis is briefly covered in this section. The study describes the different topics covered in this thesis. The study is drafted into seven chapters.

The introduction chapter, consists of the introduction and back ground information of the study including the motivation behind the study. This chapter also states the statement of the problem with their respective research questions that will be addressed at the end of the study. In this chapter we briefly stated the objective, scope, significance and limitation of the study of the proposed model.

The second chapter deals with Literature and related works that have been done in adverse pregnancy outcome Prediction using deep learning approaches. In this chapter overview common adverse pregnancy outcomes and deep learning models, deep learning in health sectors and adverse pregnancy outcome Prediction are stated in detail. This chapter also deals with the related works of previously done research.

The third chapter addresses research methodologies and design issues which includes research design, adverse pregnancy outcome predictive model with their respective dataset building, dataset cleaning and preprocessing techniques. It also states the tool that are used in the study.

The fourth chapter includes architecture of the proposed adverse pregnancy outcome Prediction model architecture, data preprocessing, predictive model evaluation and hyperparameter tuning approaches are stated.

The fifth chapter presents the various experimentation and testing for the predictive model with a snippet code of the implementation and their description. The sixth chapter discusses the result and evaluation of the proposed models. The outperforming model is selected based on the experimentation findings from the predictive model performance. The chapter also discusses how the proposed model has answered the research questions stated in chapter one. The final chapter includes conclusion of thesis, future work and recommendation direction of the study.

CHAPTER TWO

2. Literature Review and Related Work

This chapter provides an overview of various surveys of academic articles, journals, conference papers, and other sources related to adverse pregnancy outcomes and deep learning to provide a brief understanding of what was done so far, and what will be done in this thesis. Related and relevant works to our research area that are done so far are also reviewed and presented at the end of this chapter.

2.1 Adverse Pregnancy Outcomes Cause and Impacts

Even though most pregnancy and childbirth are a joyful experience for most women, it sometimes ends up with pregnancy complications(Lawn JE, 2009). Adverse birth outcomes are complex results that arise from various factors, encompassing conditions such as premature birth, low birth weight, stillbirths, birth defects, and neonatal mortality. This prevalent health concern has diverse impacts on both expectant mothers and their newborns. Birth outcomes are measures of health at birth and their magnitude has dramatically decreased in the past 40 years. However; there is still a large gap between developing and developed countries (Hornstra G, 2005).

Adverse pregnancy outcomes serve as crucial metrics for assessing the effectiveness of maternal and child healthcare programs, reflecting the quality of services such as prenatal care, labor and delivery care, and medical support. Despite Ethiopia's substantial progress in reducing maternal and neonatal mortality by half since 2000, the rates remain relatively high, with maternal mortality at 412 per 100,000 live births and neonatal mortality at 420 per 100,000 live births (CSA, 2016). Adverse pregnancy outcomes can profoundly affect the health and overall well-fare of women and their families in developing countries. These various adverse outcomes encompass various issues like maternal mortality, infant mortality, child morbidity and substantial economic hardships.

Globally, about 15 million babies are born prematurely each year, from which Sub-Saharan Africa and South Asia account for more than 60% (Lee AC, 2013). Approximately 20 million infants worldwide are born with a low birth weight, with 15% of these cases happening in the sub-Saharan African region. Globally, nearly 3 million stillbirths take place each year, and a staggering 98% of these occur in less developed nations. (Lawn JE B. H., 2016). In 2017, the top five countries with the highest maternal mortality rates were Tanzania (11,000), Ethiopia (14,000), the Democratic Republic of Congo (16,000), India (with 35,000), and Nigeria (67,000 maternal s) (Ritchie., 2017).

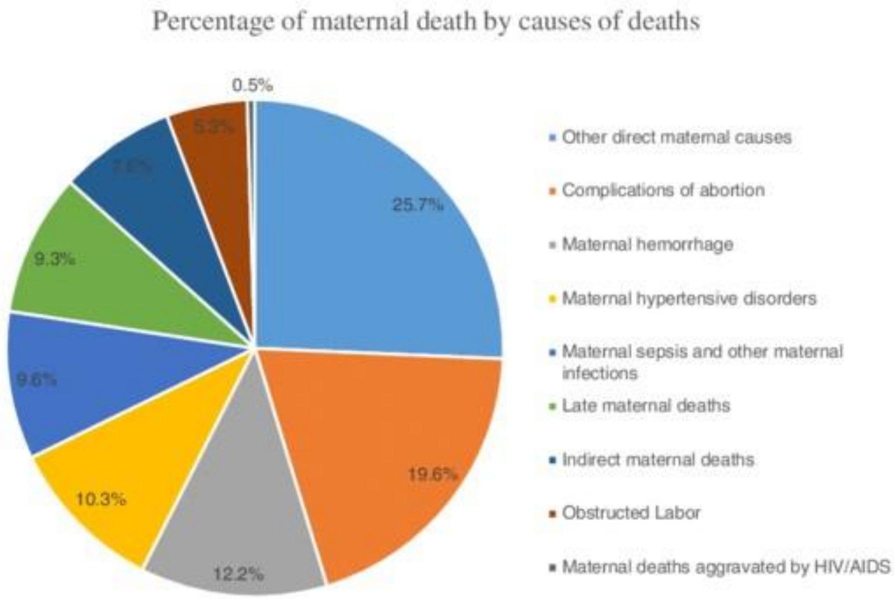

Figure 2-1 Percentage-of-Maternal-Deaths-by-Causes[5]

2.2 Common Adverse Pregnancy Outcomes

Maternal health concerns for women can emerge throughout pregnancy, labor and delivery, and the postpartum period (Finlayson K, 2020). The gestational period plays a crucial role in determining an infant's well-being and future survival. Birth weight and gestational age are two parameters that can be used to evaluate the quality of a pregnancy. The most common adverse pregnancy outcomes are the following.

[5] https://www.researchgate.net/figure/Trends-of-the-causes-of-maternal-mortality-in-Ethiopia-1990-2013

2.2.1 Intrauterine Pregnancy (Normal Pregnancy)

Intrauterine birth refers to a normal delivery in which a baby is born through the mother's birth canal after completing a full term of gestation inside the uterus[6].During intrauterine development, the baby receives all the necessary nutrients and oxygen through the placenta, which is attached to the uterine wall and provides a vital connection between the mother and the developing fetus. The duration of a normal pregnancy is around 40 weeks, and during this time, the baby undergoes significant growth and development. Intrauterine growth is closely monitored during prenatal visits, and any deviations from the expected pattern can signal potential problems with the pregnancy.

2.2.2 Preterm Birth

Preterm birth is defined as the birth of a baby before 37 weeks of gestation. Preterm birth is a significant contributor of infant mortality and can lead to a various health concern for the baby, including respiratory complications, feeding difficulties, and delays in development (Tafere Birlie Ayele, 2023). The determinant contributors for preterm birth encompass, history of prior preterm deliveries, multiple pregnancies, short cervix, maternal age, pregnancy complications, infections, lifestyle factors and Stress[7]. Complications arising from premature birth rank as the primary cause of neonatal mortality and the second most prevalent cause of death among children aged under 5 years (WHO, 2021).

In developing countries, over 60% of all deliveries are preterm, and this rate has been consistently on the rise. Despite this rate, determining the cause of premature birth in the majority of low-income countries is challenging due to a lack of precise data. The overall occurrence of preterm birth in Ethiopia was 10.48% (Muchie KF, 2020). The major contributing factor to elevated infant mortality rates, at 48 deaths per 1,000 live births, and neonatal mortality rates, at 29 deaths per 1,000 live births, is the occurrence of complications arising from preterm births (Lawn JE B. H., 2016). Preterm birth can be prevented or managed with proper prenatal care, monitoring of high-risk pregnancies, and timely interventions such as medications or bed rest. If a preterm birth is inevitable, healthcare providers can take steps to improve the baby's

[6] *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3108427/*
[7] *https://www.cdc.gov/reproductivehealth/features/premature-birth/*

chances of survival and reduce the risk of complications through neonatal intensive care and other specialized treatments.

2.2.3 Low Birth Weight

Low birth weight is characterized by a birth weight below 2,500 grams. LBW babies may be born prematurely or at full term, but they are smaller in size than normal for their gestational age. One of the risk factors associated with low birth weight is premature birth, Poor maternal health, Multiple pregnancies, Maternal age and Pregnancy complications (Girma, 2019). Low birth weight can result from intrauterine growth restriction, premature birth, or a combination of both factors. This condition leads to various adverse health outcomes, including higher rates of prenatal and infant mortality, increased morbidity, and hindered growth. Infants with low birth weight face a mortality risk of approximately 20 times greater than their heavier counterparts. Additionally, low birth weight is more prevalent in developing nations compared to developed ones. Nevertheless, a substantial number of deliveries occur in homes or small medical facilities, and information on low birth weight in developing countries is limited and often goes unreported. More than 20 million babies who weigh less than 2500 g are born each year. Globally, this statistic represents 17% of total births in low-income countries, which is more than twice the rate observed in developed countries, where it stands at 7%. (Mesfin Tadese, 2022).

2.3 Overview of Deep Learning

Deep learning falls within the realm of artificial intelligence as a subset of machine learning. Models in deep-learning, including artificial neural networks and convolutional neural networks, are capable of acquiring knowledge from intricate patterns and advanced features (Rout AR, 2018). Deep learning has a wide spectrum of uses in the domain area of medicine and health. Using deep learning techniques, the model intends to propose focuses primarily on predicting and classifying APO, which necessitates developing a model to find the early Prediction of pregnancy birth outcomes using health data.

In recent years, there has been a growing utilization of deep learning techniques to forecast unfavorable pregnancy outcomes. These methods harness the capabilities of neural networks to process extensive medical data, uncover patterns, and detect risk factors that traditional statistical models might overlook. Deep learning methods commonly applied in predicting pregnancy complications encompass Convolutional Neural Network, Recurrent Neural Networks, Long Short-Term Memory Networks, and Feedforward Neural Networks (Tobore I, 2019).

2.3.1 Convolutional Neural Network

A Convolutional Neural Network is a type of supervised deep learning model primarily employed in tasks related to image analysis. CNN includes three layers namely, convolutional layers, pooling layers, and fully connected layers. In the convolutional layer, the input image is subjected to a series of operations using kernels or filters, resulting in the creation of feature maps. In the pooling layer, the dimensions of these feature maps are decreased in order to minimize the overall number of parameters. This procedure is alternatively referred to as down sampling or subsampling (S. Chen, 2014).

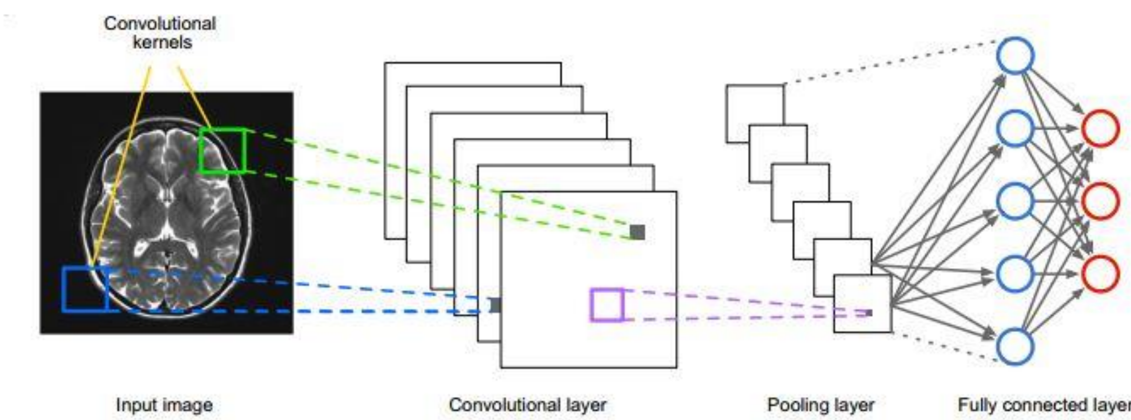

Figure 2-2 Convolutional Neural Network[8]

2.3.2 Recurrent Neural Network

Recurrent Neural Network is utilized for pattern recognition for stream or sequential data such as speech, handwriting, and text. The RNN's structure contains a cyclic connection. RNNs employ recurrent computations for sequential input processing. They store a state vector in hidden units, which retains data from prior inputs, and combine this historical information with

[8] https://mriquestions.com/deep-network-types.html

the current input to produce the next output. Moreover, RNNs consider both the current and past inputs to compute the updated output. (S. Min, 2017).

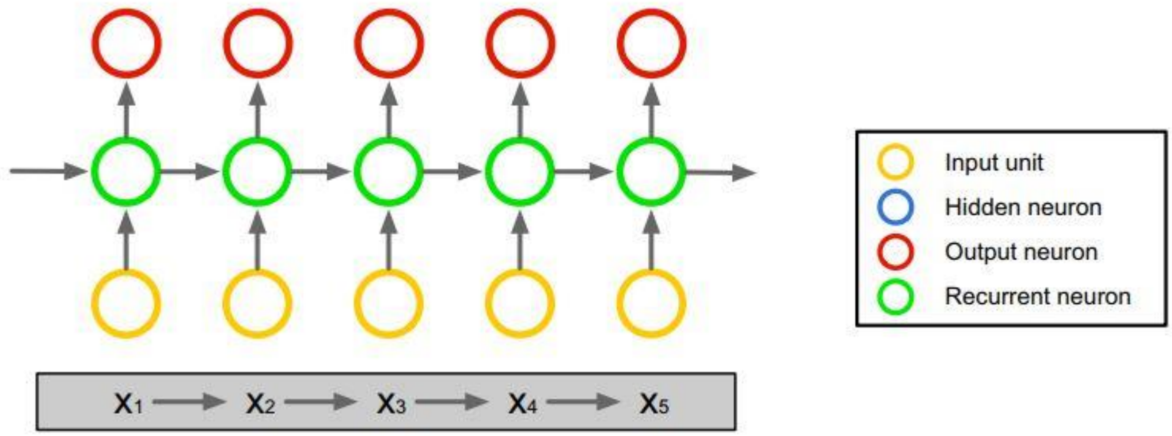

Figure 2-3 Recurrent Neural Network[9]

2.3.3 Long Short-Term Memory

Long Short-Term Memory are a type of RNN that are designed to handle long-term dependencies in sequential data. They use memory cells to store information and gates to regulate the flow of information, enabling them to make selective decision about retaining and forgetting previous inputs. LSTMs are engineered to maintain information accumulated over time, making them particularly valuable in forecasting time series data. Their capacity to retain memory of previous inputs is a key feature that contributes to their effectiveness in this context. This analogy is rooted in their chain-like structure comprising four interconnected layers that communicate with each other in distinct ways. Besides applications of time series prediction, they can be used to construct speech recognizers, development in pharmaceuticals, and composition of music loops as well(W. Sun, 2017).

[9] https://www.researchgate.net/figure/Recurrent-Neural-Network-RNN

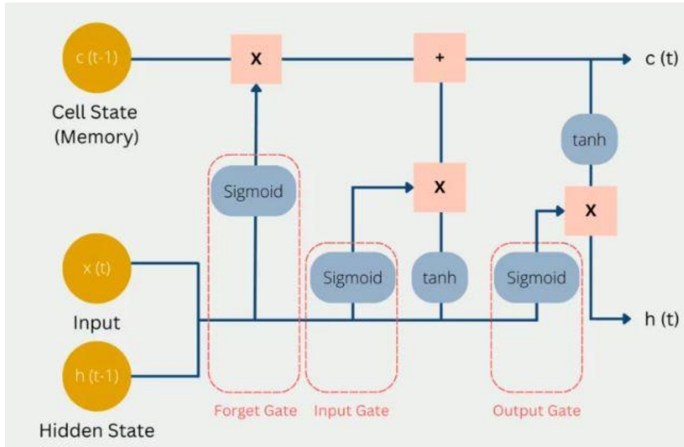

Figure 2-4 Long Short-Term Memory [10]

2.3.4 Auto-Encoder

An Auto-Encoder is a type of artificial neural network designed to encode data efficiently, making it useful for tasks like feature reduction or initializing other neural networks. It operates by processing input data through a connected neural network. An extension of this concept is the Denoising Auto-Encoder (DAE), primarily used for extracting features from noisy datasets. A typical DAE consists of three layers including input layer, encoding layer, and a decoding layer. Autoencoder Networks are used for unsupervised learning tasks such as data compression and feature extraction. These systems consist of an encoder network responsible for transforming input data into a compressed form and a decoder network that restores the initial input using this condensed form (L. Macías-García, 2017).

[10] https://what-is-the-architecture-behind-the-keras-lstm-cell

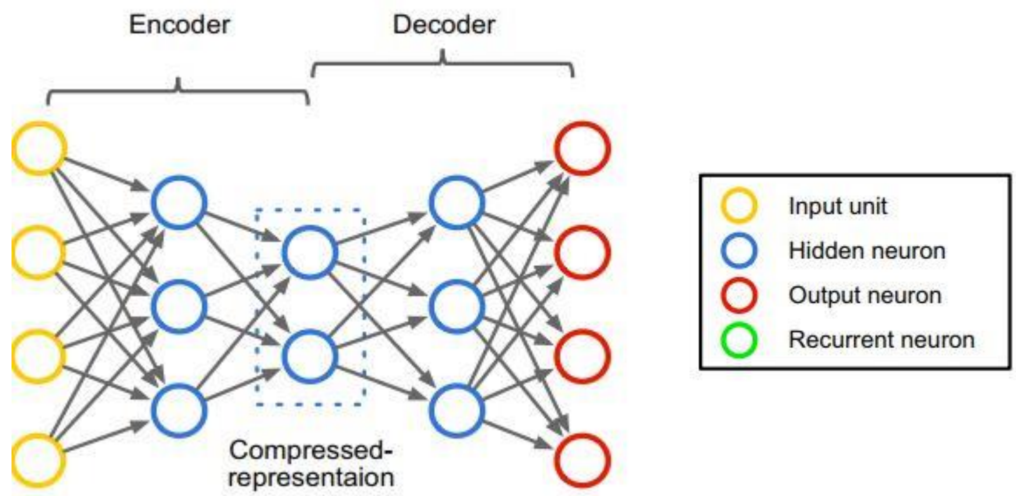

Figure 2-5 Auto Encoders[11]

2.3.5 Bidirectional Long-Short Term Memory

Bidirectional long short-term memory is a technique that enables a neural network to consider sequence information in both directions, either from the past to the future (forward) or from the future to the past (backward). Unlike a standard LSTM, a Bi-LSTM processes input data in both directions, allowing it to capture contextual information from both past and future inputs. Using LSTM, input flows in one direction, either backwards or forward. Nevertheless, in a bidirectional approach, we have the capability to allow input information to flow in both directions, ensuring that both past and future information are retained (Islam, 2018).

[11] https://neptune.ai/representation-learning-with-autoencoder

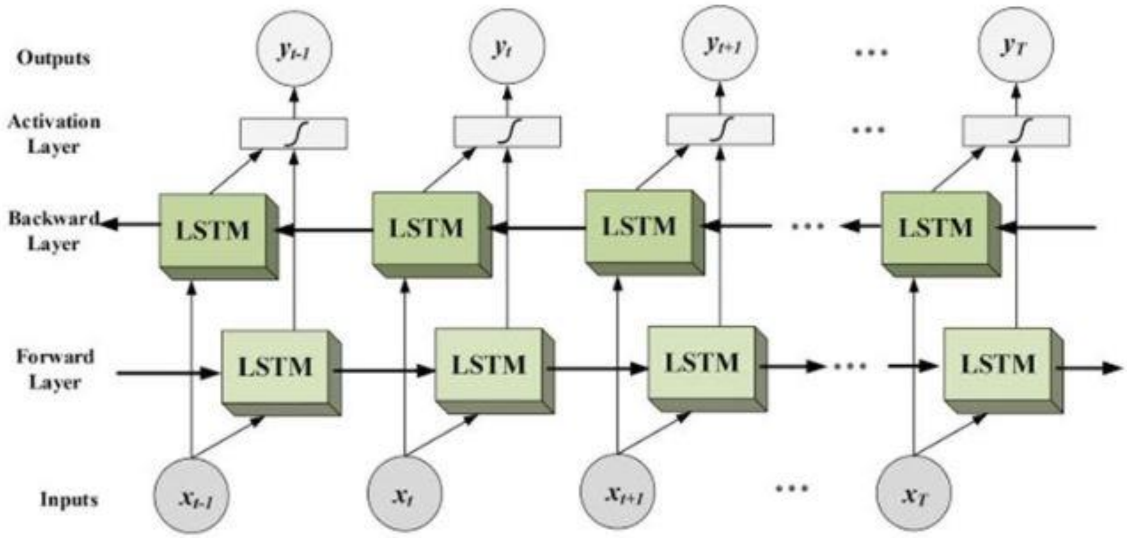

Figure 2-6 Bidirectional Long Short-Term Memory[12]

2.3.6 Feedforward Neural Network

The feedforward neural network represents one of the most straightforward artificial neural network architectures. It arranges neurons in sequential layers, with the initial layer serving as the input layer, where each unit corresponds to a specific dimension of the data vector. The output layer is the last layer in the network, responsible for indicating the probability of an object belonging to various Prediction categories. Hidden layers, on the other hand, are layers situated between the input and output layers. (Chang Su, 2020).

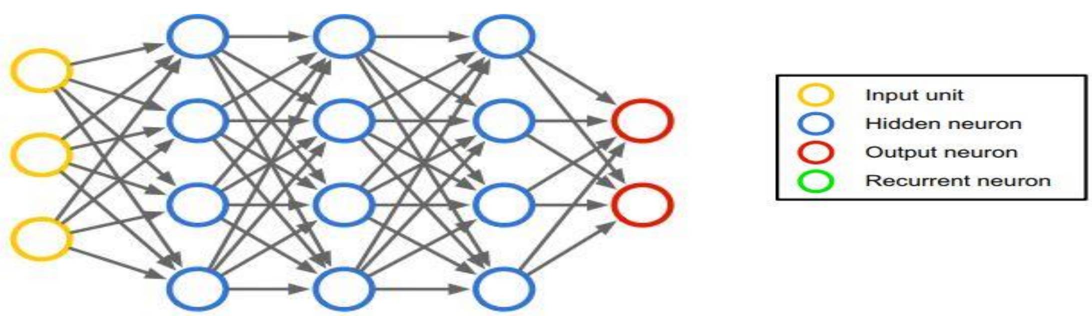

Figure 2-7 Feedforward Neural Network[13]

[12] https://analyticsindiamag.com/complete-guide-to-bidirectional-lstm-with-python-codes/
[13] https://www.researchgate.net/figure/Feedforward-Neural-Network-FNN

2.4 Deep Learning in Health Sector

In recent times, the healthcare sector has witnessed the adoption of big data technology, driven by the exponential growth in biomedical data and the rapid progress in both medical and computer technologies. The broad phrase healthcare encompasses the preservation and improvement of people's health through diagnosis, prevention, and treatment, as well as the recovery from or even the curing of illness, disease, accident, or any further physical or mental afflictions (Miotto, 2018). The significance of healthcare data analysis is increasingly prominent in the realm of personalized medicine. In the last few years, given the increasing amount and complexity of healthcare information, the use of machine learning techniques like deep neural network models has become more alluring (Shahab Shamshirband, 2021).

Deep learning techniques involve a multi-stage process of feature extraction, where data passes through a series of layers to progressively learn and represent features. Deep learning has the capacity to bring about a significant transformation of the healthcare industry by providing more accurate diagnosis, tailored treatments, and better patient outcomes. Deep learning is being used in healthcare sectors such as medical imaging, electronic health data, drug discovery, personalized medicine, disease outbreak prediction and medical robotics. The development and deployment of these technologies must be done ethically and responsibly, taking into account concerns like data privacy, regulatory compliance, bias and informed consent.

2.5 Deep Learning Approach in Adverse Pregnancy Outcome

In the present era, individuals encounter a multitude of health issues as a result of both environmental factors and their lifestyle choices. Therefore, detecting diseases at an early stage is a crucial and significant task (Dhiraj Dahiwade, 2019). Adverse pregnancy outcomes are a serious public health concern around the world, and identifying women at risk of pregnancy complications is crucial for improving maternal and fetal health outcomes. Deep learning can improve APO prediction by examining vast datasets of maternal and fetal health data.

The number of maternal fatalities associated with pregnancy among women per 100,000 live births is known as the maternal mortality ratio (WHO, 2019) . Medical conditions that affect pregnant women, such as age and blood diseases, changes in vital signs like blood pressure, blood sugar levels, body temperature, and heart rate can directly escalate. The number of issues that happen throughout pregnancy, which can lead to miscarriage and the death of both the expectant mother and the infant. These health conditions require the use of specialized medications prescribed by medical professionals, and early identification of these risks may enable the professionals to take the necessary preventative measures to lower the risk of mortality. These types of deep learning approaches are utilized in healthcare systems to describe the presence of adverse and severe health problems in an individual: -

Table 2-1Pros and Cons of Deep Learning Algorithms Used in Maternal Health Care

DL Algorithm[14]	Pros	Cons
CNN	Achieving high precision in image recognition tasks through weight sharing.	A substantial amount of training data is required, Do not include the information regarding the object's position and orientation in the encoding.
RNN	Handle any length of the input model size, provide the model with context for the entire text, useful in time series prediction.	Training RNN is a very difficult task, Gradient vanishing and exploding problems
BiLSTM	Utilizes information from both sides, produce more meaningful output, combines LSTM layers from both directions	Much slower model, requires more time for training,
DBN	Has specific robustness in Prediction, avoids time taking techniques.	Expensive to train because of complex data models, difficult to be used by less skilled people.
FNN	Efficient training, simplicity, universal approximation.	Limited input flexibility, lack of interoperability.
LSTM	Better at handling long-term dependencies, less susceptible to the vanishing gradient problem and very efficient at modeling complex sequential data.	Require more training data to learn effectively. LSTMs may not be appropriate for all types of data

[14] https://www.geeksforgeeks.org/advantages-and-disadvantages-of-deep-learning/

AE	Reduces requirement for feature engineering, eliminates unnecessary expenses, effortlessly recognizes challenging flaws and delivers top-tier performance on complex issues."	Necessitates a substantial quantity of data, quite expensive to train does not have strong theoretical groundwork.

2.6 Related Works in Adverse Pregnancy Outcomes

(Muhammad Nazrul Islam, 2022) This research delved into the latest perspectives on research and development in the field of machine learning, with a specific emphasis on its applications for forecasting and predicting various pregnancy-related conditions. The research presented a methodology for implementing machine learning algorithms to improve maternal healthcare in the future by identifying preventable complications during pregnancy. Deep learning and ensemble machine learning can be utilized in the future due to the substantial performance enhancements they bring to the field of medical diagnosis.

(Hasan Rawashdeh, 2019)The study projected to develop in two steps. The primary objectives were to create two distinct models: one to serve as a decision support system for pregnant women at elevated risk of premature delivery before undergoing cervical cerclage, and the other to forecast the timing of spontaneous delivery in this high-risk group following cerclage. To construct the prediction models, various techniques such as SMOTE, Decision Tree, Random Forest, K-Nearest Neighbors, and Neural Network were employed. Among these methods, the Random Forest classifier yielded the most favorable outcomes, demonstrating a Glean score of 0.96 and sensitivity of 1.00. The research paper introduced computational models that could be employed to anticipate the necessity for cerclage and predict the gestational age at delivery subsequent to this procedure.

(Belayneh Endalamaw D, 2022) This paper projected to predict the level of anemia among pregnant woman in Ethiopia using homogenous ensemble machine learning. The researchers used Decision tree, Random Forest, Cat boost, and extreme gradient boosting method for prediction. Health data is collected from an Ethiopian demographic health survey and preprocessed to get data that are suitable for Machine Learning. The total accuracy of random forest, extreme gradient boosting, and cat boosts with one versus the rest are 94.4%, 94.54%, and 97.6%, respectively.

(Cheng Gao, 2019)This paper projected to investigate the extent to which deep learning method documented in HER can predict EPB. Word embedding and Bag of Words, Cohort construction, which uses bootstrapping, Traditional machine learning models (LR, SVM, GB) and deep learning model (RNN) and LSTM) were used to predict EPB. The RNN ensemble models outperformed the RNN models trained on naturally imbalanced EPB ratio datasets in terms of both AUC (0.827 compared to 0.744) and sensitivity (0.965 compared to 0.682).

(Ayleen Bertini, 2022) The objective of this research was to assess the effectiveness and applicability of machine learning techniques for detecting pregnancy complications. Tree-based Pipeline Optimization Tools and Convolutional neural network were used to predict pregnancy complications. The MLM with the most successful outcomes was the prediction of preterm from medical pictures using the SVM approach, which had an accuracy of 95.7%, and the prediction of newborn death using the XGBoost technique, which had an accuracy of 99.7%.

(Yandamur, 2018) The researchers utilized a multilayer perceptron in their effort to create a clinical decision support system aimed at forecasting pregnancy results for expectant mothers diagnosed with Systemic Lupus Erythematosus (SLE). RNN with LSTM was used to predict the pregnancy outcome with the help of MLP. To predict the outcomes of pregnancies in women with SLE, a clinical decision support system based on RNN is developed. The overall accuracy of RNN with LSTM accuracy, precision, and recall are 96.7%, 97%, and 98%, respectively.

(Hisham Allahem, 2022). This study used machine learning and deep learning techniques to anticipate and predict labor in order to mitigate the impact of premature birth on pregnant women and the fetus. The study proposed an approach to automatically predict and predict labor by classifying EHG signals using ML and DL algorithms. The deep learning approach achieved the best result with a 0.98 % accuracy rate.

(Melissa J Fazzari, 2022). This study aimed to investigate whether machine learning methods might better predict APO and identify further risk factors. (LR-S), (LASSO), (RF), (SVMRBF), gradient boost GB), neural networks and Super Learner were used to predict APO in patients with SLE. SL, which combines predictions from four distinct algorithms, exhibited the highest overall capability to distinguish between APO and non-APO cases, achieving an Area Under the Curve value of 0.78.

In general, all the above-mentioned papers have their contributions to their respective field and areas. Apart from that there are gaps encountered in their paper. A study conducted by (Yu Mu, 2018) has used a large number of feature set that makes feature engineering and selection challenging. Because managing and selecting relevant features from a large pool can be difficult. These affected the performance of the model negatively. In addition, optimization techniques were not incorporated to the fullest into it. (Dithy M. D, 2022) They have used a limited number of datasets and the results might be biased. (Yandamur, 2018) In this paper the predictive accuracy of CDSS for pregnancy outcomes before conception is limited, and only a small set of data has been utilized for such predictions. (Yu Mu, 2018) This paper needs to apply feature engineering and optimize the incidence of adverse pregnancy outcomes. (Cheng Gao, 2019) Because they used data from a single health organization, results may be biased, the root cause of poor PPV is not mentioned and Refined models are needed to learn the patterns differentiating control deliveries. (Hasan Rawashdeh, 2019) In this paper additional new instances to the model needs to be added in order to enhance the metrics of prediction accuracy. (Hisham Allahem, 2022) This paper needs to incorporate with additional parameters in order to yield the preferable to(Ayleen Bertini, 2022) Sample size is not representative of the given geographical area. It is difficult to compare the prediction accessed because they have, used different baselines, variable input and separate complications and some of the predictions lack variable measurements.

This research will focus on using deep learning model to solve the problem that's has been encountered in a national level by using predictive model for adverse pregnancy outcomes. As a result, this research will incorporate a wider range of models to enhance the precision of predicting adverse pregnancy outcomes occurrence by using EDHS health data. In this study, deep learning algorithms will be used for a more in-depth analysis. The overall outcome of the study is to improve pregnancy outcomes, reduce the incidence of prematurity and LBW and finally reduce maternal and infant mortality and morbidity rate.

Table 2-2 Summary of Related Works

Author and Year	*Objective*	*Best Algorithm*	*Gap*
(Yandamur, 2018)	To anticipate the pregnancy results of women who have Systemic Lupus Erythematosus during their pregnancy.	LSTM	The study focused only on SLE-affected pregnant women. They have not described the features. They have used a small number of datasets. Most contributing attributes were not included. It is a binary Prediction problem.
(Cheng Gao, 2019)	To investigate the extent to which DLM documented in EHR can predict EPB.	RNN-LSTM	Due to the fact that they have used data from a single health organization, results may be biased and reasons for Poor Preterm delivery reasons are not clearly stated.
(Hisham Allahem, 2022)	To reduce the impact of premature birth for pregnant women and fetuses using ML and DL	DL	Used only one feature to address the problem. Important evaluation metrics were not included. It is a binary Prediction and they didn't develop a predictive model. The number of datasets is small and they need to incorporate a large data size.
(Ayleen Bertini, 2022)	To identify the applicable performance of ML approaches to predict pregnancy outcomes	SVM	Some predictions lack variable measurements and it is difficult to compare the prediction accessed because they have used different baselines, variable input, and separate complications.

(Yu Mu, 2018)	To use the DLM to predict adverse pregnancy outcome using a Multilayer Neural Network	MLNN	Optimization techniques were not obtained to the fullest. More feature engineering techniques should be applied.
(Lakshmi.B.N, 2015)	The study aims at monitoring women during pregnancy to protect them from possible pregnancy complications.	C4.5	Incorporate with large dataset. Additional parameters and methods should be incorporated to it.

Based on the gaps mentioned above, this study comes up with a solution for the identified problems that will benefit the expectant mother, fetus, health care professionals' challenges by predicting early presence of adverse pregnancy outcome using deep learning approaches. The outcomes of this study will reduce mortality and morbidity rate by allowing health care professionals intervene with their expertise for the expectant mother and families.

CHAPTER THREE

3. Research Methodology

3.1 Overview of Research Methodology

The chapter's main objective is the construction of the methodology. The methods and processes employed to accomplish the goal of the current study are described in this section. This chapter discusses the overall methodology of developing adverse pregnancy outcome Prediction model based on the determinant risk factors using deep learning techniques. Deep learning has a several range of potential applications in the field of medicine and health. Different available approaches and evaluation metrics used for the performance are also provided in this chapter. More over to build the predictive model different data collection, preprocessing, evaluation metrics and deep learning algorithms techniques and methods are used and explained briefly throughout the research.

3.2 Research Design

A research design involves structuring data collection and analysis methods in a way that achieves a balance between alignment with the study's objectives, efficiency, and methodology (De Vaus, 2001). Research design is the plan strategy investigation concaved to obtain research question and control variance. They further indicate that an effective study design strives to generate dependable results.

The goal of a research design is to ensure that the information acquired allows us to clearly and rationally address the research topic. (De Vaus, 2001). There are 5 common types of research designs. These designs are descriptive, correlational, quantitative, explanatory and experimental (Jovancic, 2020.). This research implemented an experimental and design science research methodology because it aimed to establish relationships between dependent and independent variables through a quantitative research design approach.

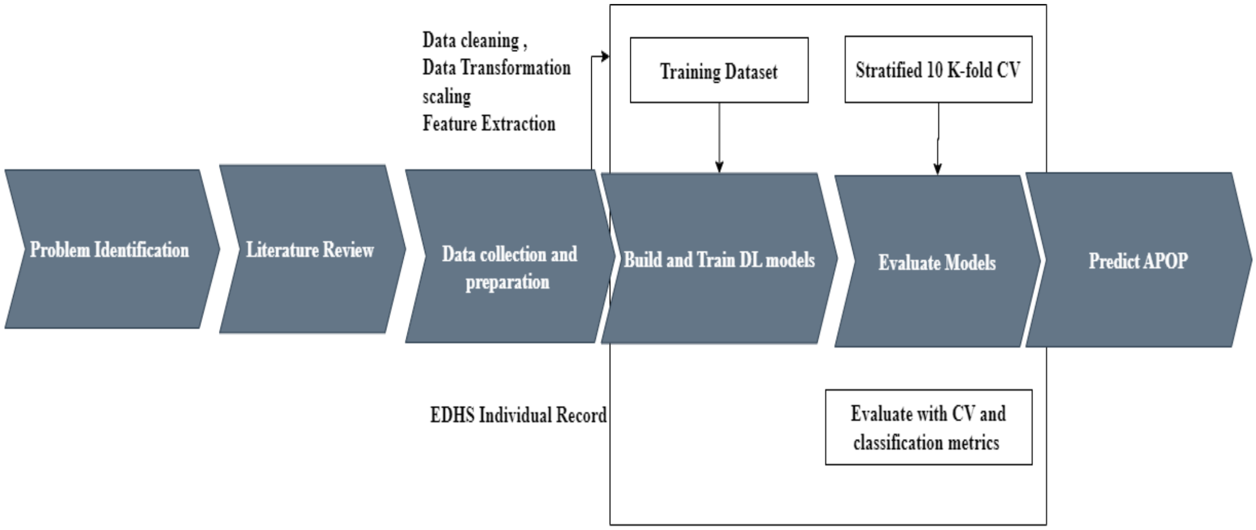

Figure 3-1 Methodology for adverse pregnancy outcome prediction

3.3 Adverse pregnancy outcome Prediction Modeling

3.3.1 Dataset Building

The main aim of this study is to predict the presence of adverse pregnancy outcome using deep learning techniques that labels the outcomes as Normal, premature birth, and low birth weight. Enhancing the quality of the data is the fundamental step in accomplishing the study's goal. The data collected from EDHS should be preprocessed and determinant risk factors should be identified by using feature selection and extraction techniques.

3.3.2 Data Collection

The data utilized in this study was gathered from the Ethiopian Demographic Health and Survey (EDHS), conducted by the Ethiopian Central Statistical Agency in the years 2000, 2005, 2011, 2016, and 2018, with each survey conducted at five-year intervals.

3.3.3 Data Preprocessing

Data preprocessing involves essential initial step in developing a deep learning model, which is preparing raw data to render it suitable for use. Real-world data often contains irregularities like noise, missing values, and may not be readily usable for machine learning models, requiring the need for data preprocessing.[15]. Data preprocessing is an essential step for cleansing and readying the data to be used with a deep learning model, ultimately enhancing both the accuracy and efficiency of the model. The data taken from EDHS are raw data that need to be preprocessed by using data cleaning, data integration, data transformation denoising, feature selection and data splitting techniques.

3.3.4 Data Cleaning

Data cleaning is the process of identifying and correcting or removing errors, inconsistencies, and outliers in the raw data. This is an essential step in data preprocessing, as the quality of the data is essential to building accurate and effective deep learning models. Data cleaning involves handling missing values, handling outliers, handling duplicates, handling irrelevant and inconsistent data. Data cleaning helps to guarantee that the data is consistent, complete, and relevant to the problem we are trying to solve[16].

3.3.5 Handling Missing Values

Handling missing values is an important step in data preprocessing, as missing data can cause issues with deep learning algorithms. In real-world datasets, the presence of missing values is a frequent issue, and these gaps in data can arise for a variety of reasons, such as measurement errors, data corruption, or human error during data entry. Handling missing values involves filling in the missing data with an appropriate value or removing the rows or columns with missing values[17]. The approach to address missing values is contingent upon both the data type and the specific problem being addressed.

3.3.6 Handling Categorial Data

Handling categorial data is an important step in data preprocessing, as deep learning algorithms typically work with numerical data[18]. The data that is collected from EDHS dataset contains nominal, real and decimal data that needs to be changed in to numerical to feed the data in to deep learning model. In individual women's record "V363" Preferred future method for

---

[15] *https://www.techtarget.com/searchdatamanagement/definition/data-preprocessing*
[16] *https://www.tableau.com/learn/articles/what-is-data cleaning*
[17] *https://www.analyticsvidhya.com/blog/2021/10/handling-missing-value/*
[18] *https://towardsdatascience.com/handling-categorical-data-the-right-way*

respondents intending to use a method in the future having 2 possible outcomes (V312 = 0 & (V362 = 1 or V362 = 2 or V362 = 3)). these values stand for 0= “not currently using contraception”, 1= “unsure about use”, 2= “unsure about timing”.

3.3.7 Handling Imbalanced Data

Handling imbalanced data refers to datasets in which there is an unequal distribution of classes, with one class having notably more instances than the others. Handling imbalanced data is a vital component in data preprocessing, as machine learning algorithms can be biased towards the majority class when dealing with imbalanced datasets. By using techniques such as oversampling, under sampling, or both we can address class imbalance. Oversampling refers to a method of addressing class imbalance by boosting the representation of the minority class, achieved either through duplicating existing instances or generating synthetic data. Conversely, under sampling entails reducing the number of instances in the majority class by randomly removing some of them to balance the dataset. To avoid loss of valuable information for the training data sets and to treat the imbalanced data SMOTE and stratified K-Fold cross-validation with conjunction will be used for this study. While stratified K-Fold cross-validation helps in distributing the data evenly across different folds to ensure that each fold maintains the class distribution of the original dataset, it may not completely address the issue of imbalanced data. Using SMOTE in conjunction with stratified K-Fold cross-validation is a common strategy to tackle class imbalance by learning imbalanced patterns, improving the generalization of the model, mitigating the bias, and ensuring the deep learning model performs well across different folds in the presence of imbalanced data. It's a complementary technique that can enhance the robustness and effectiveness of the model, especially when dealing with datasets where one class is significantly less frequent than the others.

3.3.8 Feature Extraction

Feature extraction is a technique for reducing dimensionality that partitions a large initial dataset into smaller subsets, facilitating more efficient data processing. It is also the process of choosing and combining variables into features. This method produces high-quality results while reducing the amount of data that needs to be processed. Feature extraction can help reduce the volume of repetitive or unnecessary data during an analysis (Jogin, 2018). For this study we have used MLP feature extraction method because of the nature of our dataset. Fully connected neural networks consists of one or more hidden layer to extract the higher-layer representation of patterns from the input feature. The majority of scholars use these MLPs because they provide flexibility and generalizability to learn a mapping from inputs and outputs. They are suitable for prediction problems where input is assigned as class and label (Dong, 2021). Because of the nature of our dataset being tabular and having prediction problems, MLP is the best suite for feature extraction. One of the benefits of MLP having multiple layers allows the model to learn complex representation and capture non-linear relationships of the dataset.

3.3.9 Feature Selection

Feature selection involves the act of picking out a more focused set of important attributes (or variables) from a larger pool of available features within a dataset. The main objective of feature selection is to enhance the performance of a machine learning model by reducing the complexity of the data, and by eliminating inconsistent, duplicated, or noisy features that could potentially harm the model's accuracy, comprehensibility, or ability to make generalized predictions. The goal of feature selection techniques in machine learning is to find the perfect combination of features that make it possible to generate optimum models of researched phenomena.

3.4 Deep Learning Models

Deep Learning is a subfield of Machine Learning that employs multiple layers of data processing to enable machines to acquire knowledge from data by progressively abstracting information at various levels. It has gained more attention for NLP tasks, Maternal related tasks from the ML research communities (Alshahrani, 2016). Selecting the best possible deep learning approach for a specific task requires considerations of multiple criteria.

The first criterion is understanding the specific task requirement, especially the nature of the problem whether they are Prediction, prediction or regression. The problem we are going to address in adverse pregnancy outcome Prediction using deep learning approach is a multi-class Prediction problem. The dataset we have collected from EDHS was created as CSV with multiple features having a target variable to which the class Prediction belongs. Deep learning has been overwhelmingly effective for multi-class Prediction problem (Junyi Chai, 2021).

The second criterion is model architecture. Different deep learning approaches has various model architecture that are well suited to specific task and data types. The nature of the dataset we have used for study was a labeled dataset that have a known target variable which comes to the category of supervised learning. By doing so we can narrow our search in to a deep learning approach that are well suited for adverse pregnancy outcome Prediction by using a labeled preprocessed dataset.

The third criterion is the most recent, widely used and outperforming deep learning approach from previous works on maternal health care. These algorithms are FNN, LSTM, BiLSTM. The final criterion is experimentation and hyperparameter tunning on the preprocessed dataset to get the outperforming deep learning model. Several experimentations will allow us to make a final decision on the best suited model for adverse pregnancy outcome prediction. For this study we have proposed the following deep learning algorithms

3.4.1 LSTMs

LSTMs are designed mainly for time series predictions because they can hold onto memories or prior inputs since they are intended to be maintained throughout time. LSTM is the most widely utilized type of recurrent neural network. LSTM network is comprised of three distinct gates, each serving a specific function.

Forget Gate: When the sigmoid activation function is used, the resulting output will fall within the range of 0 to 1. This output signifies whether the information is being discarded (0) or retained (1).

Input Gate: It is in charge of conveying the significance of any new information contained in the input. Here, we shall use the sigmoid and tanh activation functions. The functionality of a sigmoid is to retain or discard information, whereas the functionality of a tan is to either remove information from the cell state or add new information to the cell state.

Output Gate: The statement we provide is provided to the output gate as a fill-in-the-blank so that it can determine the outcome based on the information it perceives.[19]

[19] *https://www.geeksforgeeks.org/long-short-term-memory-lstm-rnn-in-tensorflow/*

3.4.2 Feedforward Neural Network

Feedforward artificial neural network is a deep neural network model with a number of completely connected layer[20]. Modern deep learning models require significant computational resources, but one can leverage MLP, as an alternative to mitigate these resource demands. Each successive layer consists of nonlinear functions applied to the weighted sum of all the outputs from the previous layer, and these connections are fully interconnected.

3.4.3 BiLSTM

BiLSTM calculates the input sequence from the reverse direction, resulting in the computation of a forward hidden sequence and a backward hidden sequence. The final forward and backward outputs are combined to create the encoded vector. Bidirectional recurrent neural networks are employed to overcome the constraint of a single LSTM cell, which can solely capture past context information and cannot leverage information from future context. Bidirectional recurrent neural networks employ two distinct LSTM layers that process input data in both forward and backward directions, ultimately converging on the same output. (Jour Schwenker, 2019).

3.5 Model Evaluation

To construct the model that predicts the presence of adverse pregnancy outcome in case of Ethiopian women's EDHS dataset was used by training it over deep learning models. For developing the predictive models, determinant risk factors are selected by the step forward feature selection method and three domain experts (2 MSc and 1 BSc holders) who are affiliated at EPHI and Tikur Anbesa referral hospital professionals were used.

3.6 Predictive Model Evaluation

Model evaluation involves the use of various metrics to assess a machine learning model's performance, identifying its capabilities and limitations. It serves a crucial role in gauging a model's effectiveness during the early stages of development and is also essential for ongoing model monitoring. Thus, the following are the evaluation metrics used to evaluate the proposed model.

---

[20] *https://towardsdatascience.com/multilayer-perceptron-explained*

3.6.1 Cross-Validation

Cross-validation is a technique where the entire dataset is divided into smaller subsets, and then the deep learning model is assessed using different subsets of data to gauge its accuracy. This approach is commonly employed in model evaluation to prevent overfitting and enhance the model's ability to generalize effectively (Sashikanta Prusty, 2022).

3.6.2 K-Fold Cross Validation

The variable 'k' is used to specify the number of segments into which the dataset will be split during K-Fold Cross-Validation. Consequently, this means that every fold within the dataset can contribute (k-1) times to the training set, guaranteeing the inclusion of all perspectives in the dataset and enabling the model to more effectively gauge its actual performance. The major benefit of using the K-Fold CV technique is that it doesn't care about specific way of the data is partitioned (Bhatt, 2021).

3.6.3 Stratified K-fold Cross-Validation

An enhanced version of K-Fold Cross-Validation ensures that each fold in the dataset maintains a consistent proportion of data points for every class label. This variation is particularly useful for addressing Prediction tasks with imbalanced class distributions. The cross-validation tool generates these balanced folds and is referred to as a stratified version of K-Fold. It involves reorganizing the data so that each fold or group is a good representation of the entire dataset. To deal with the bias and variance, it is one of the best approaches (Allen, 2021). Due to the presence of imbalanced dataset, we have used this method for our study because the stratified k-fold ensures that each fold represents the complete data.

3.6.4 Prediction Metrics

The most commonly used metrics for evaluating Prediction performance are accuracy, precision, and the confusion matrix.[21]

Accuracy represents the count of correct predictions made by the predictive model among all the predictions it has generated. This value is determined through the following calculation. It is calculated as:

[21] *https://www.dominodatalab.com/data-science-dictionary/model-evaluation*

$$\frac{TPR + TNR}{TPR + FPR + TNR + FNR} \tag{1}$$

while TPR, represents the proportion of correctly identified positive samples by our predictive model. The FPR, indicates the proportion of incorrectly identified positive samples. TNR, signifies the ratio of accurately predicted negative samples, while FNR, reflects ratio of negatively identified samples.

Precision is a measure that indicates the proportion of accurate predictions made by the model among all the expected positive data samples. It is calculated using:

$$\frac{TPR}{TPR + FPR} \tag{2}$$

Recall shows the preciseness of the model and is calculated using:

$$\frac{TPR}{TPR + FNR} \tag{3}$$

F1-Score performance metric is based on the combination of recall and precision and is calculated by taking the harmonic mean of recall and precision.

$$\frac{2 * \text{precision} \ * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

3.6.5 Confusion Matrix

A confusion matrix is a technique used to consolidate the evaluation of how well a model performs in classifying data, which can involve multiple classes within a test dataset. It organizes Prediction results into a matrix format. This matrix is structured so that one dimension represents the actual or true class of an object, while the other dimension represents the class assigned by the classifier. The matrix is a summary representation of prediction outcomes, indicating the total count of both accurate and erroneous predictions. The matrix looks like as below table that was used for three Predictions in this study for adverse pregnancy outcome were classified as Normal Birth, Low Birth Weight and Preterm Birth

Table 3-1 Confusion Matrix Predictive Values

		Predicted Values		
		Normal Birth	***Low Birth Weight***	***Preterm Birth***
Actual Values	***NB***	True NB	False LBW	False PTB
	LBW	False NB	True LBW	False PTB
	PTB	False NB	False LBW	True PTB

CHAPTER FOUR

4. Proposed Adverse Pregnancy Outcome Prediction

In this chapter, a detailed discussion was presented on the proposed Adverse pregnancy outcome Prediction model that utilizes deep learning techniques to classify adverse pregnancy outcomes. This model aims to address the issues highlighted in chapter one and provide insights into the research questions posed, thus bridging the disparity highlighted in the literature review. Adverse pregnancy outcome Prediction model architecture is thoroughly described in this section. The models key components, their respective subcomponents and how they interact with each other will be visually represented.

4.1 Proposed Adverse pregnancy outcome Prediction Architecture

The main aim of this study is to develop a deep-learning based model for predicting adverse pregnancy outcome, resulting in a significant improvement in the predictive model accuracy. The model aims to classify adverse pregnancy outcomes such as low birth weight (LBW), Preterm Birth (PTB) and Normal Birth based on maternal health care dataset obtained from EDHS. Deep learning techniques are employed to develop a model for predicting the prevalence of pregnancy-related complications in pregnant women, specifically in the context of Ethiopia.

The overall methodology will be undertaken with in preprocessing, algorithm and training parts by using different techniques and evaluation metrics to enhance the overall data quality. The data gathered from EDHS are raw data. To use the dataset for the deep learning algorithm, the proposed architecture specifies the need to preprocess EDHS dataset using data cleaning, data integration, data transformation denoising, feature selection and handling categorical values. Cross-validation techniques are utilized in this step to find appropriate feature selection techniques to assess the strength of the association between target variables, and features. The proposed dataset will be fed to deep learning models with feature selection and engineering methods. Finally, the effectiveness of each predictive model was assessed using metrics such as accuracy, precision, recall, and the F1-score.

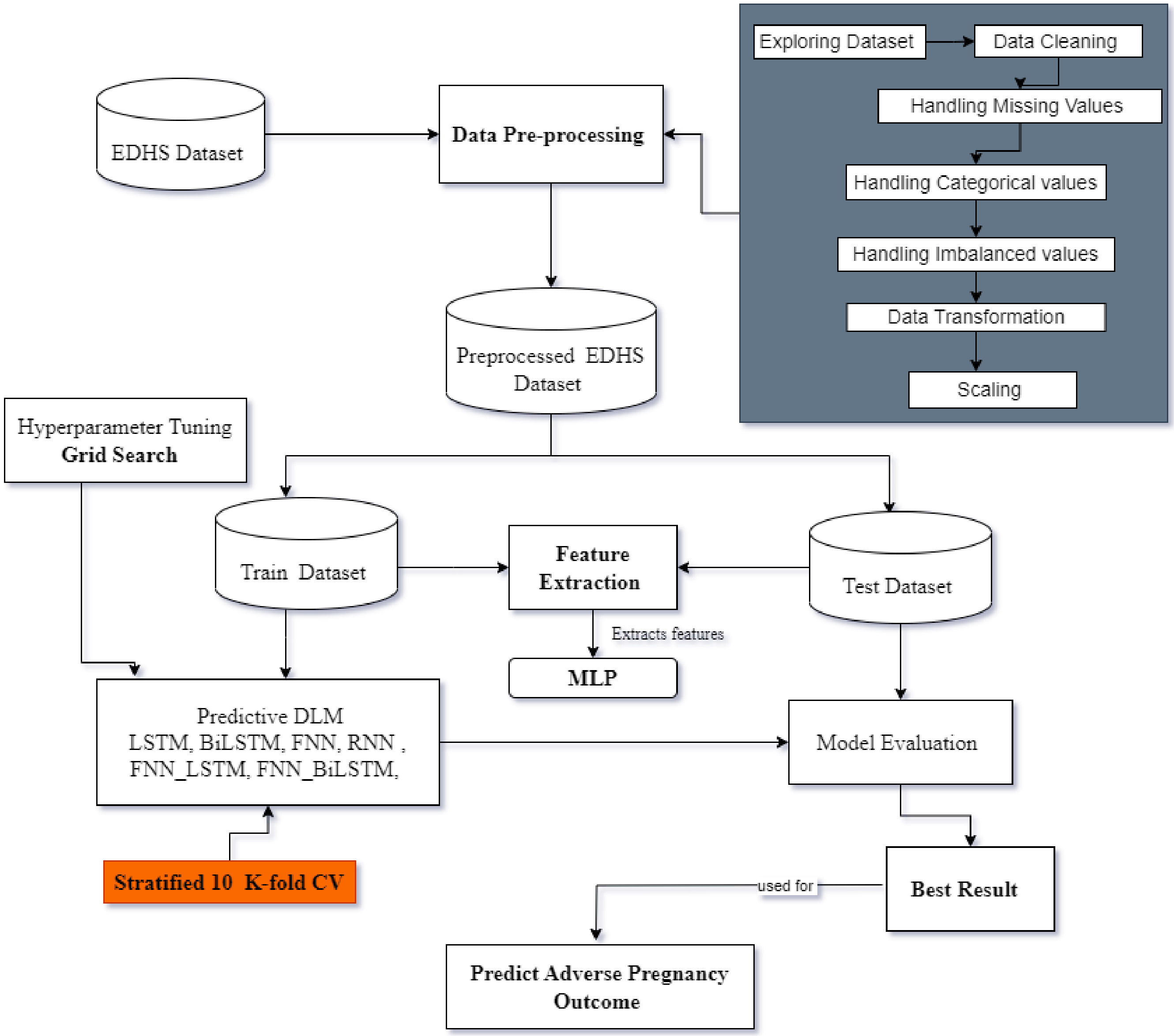

Figure 4-1 Proposed Architecture of APOP

4.2 Understanding of the Data

Data is a significant element that supports model training, evaluation, and generalization in the field of deep learning rather than merely being a collection of inputs and outputs. After identifying the problem, the next phase is to understand the dataset that is being collected from EDHS. These phases include understanding the basic characteristics of the dataset, including the number of samples, features, explore their data types, distributions, ranges, and statistics of each feature. Deep learning complexity can be more easily navigated and useful insights can be gained by having a greater grasp of the data. Finally, the dataset is checked for suitable data transformations, such as addressing missing values, outliers, or class imbalances, to make sure the data is well-prepared for training deep learning models. Considering the comprehensive understanding of the dataset, the effectiveness and completeness of the dataset concerns deep learning approaches to be used for adverse pregnancy outcome Prediction.

The dataset utilized in this study comprises individual records (IR) extracted from the Ethiopian Demographic and Health Survey conducted at five-year intervals, spanning the years 2005, 2011, 2016, and 2018(mini-EDHS) with instances of 14070, 16515, 15680, and 8885 respectively. These records were meticulously processed to ensure data quality and consistency. The data processing techniques encompassed handling missing values, data imputation, feature engineering, and normalization. In total, 10,220 instances, each characterized by 15 features with one target variable including the 3 class labels, wereselected for further analysis in the development of a Prediction model. These features encompass a range of variables relevant to the research question, including socio-demographic, obstetric history and characteristics, lifestyle and environmental factors, immunization, domestic violence, birth history, history of disease and drug use and medical history. This study aims to harness this dataset to address a specific research objective, underscoring the significance of the data selection and preprocessing steps in ensuring the dataset's suitability for model development. They offer comprehensive insights into the demographic and health-related conditions of the Ethiopian population, enabling evidence-based decision-making and the design of targeted interventions to improve healthcare and well-being in Ethiopia.

Folic Acid and Iron Supplement [M45$1]: - Folic acid supplements are a type of Vitamin B9 used to address deficiencies and reduce the risk of pregnancy-related complications. Numerous dietary sources naturally contain folate.[22]

Pregnancy complication [V228]: Complications during pregnancy encompass both physical and mental conditions that impact the well-being of the pregnant individual, the baby, or both

[22] *https://www.webmd.com/vitamins/ai/ingredientmono-1017/folic-acid*

during the pregnancy or postpartum period. Physical and psychological issues that can be problematic can appear throughout pregnancy.[23]

Size of the child [M18$1]: - Birth weight is a crucial health metric for infants. Full-term infants typically (born between 37- and 41-weeks' gestation) weigh 3.2 kg on average. Generally, infants with low birth weights and those born significantly above average weight are at a higher risk of experiencing complications.[24]

Weight [V437]: - refers to the body weight of a woman during pregnancy. Being overweight and underweight during pregnancy can cause complications for the expectant mother and the baby. Pregnancy often leads to weight gain due to various factors such as the growth of the fetus, increased blood volume, amniotic fluid, and changes in maternal tissue. The weight of pregnant women is monitored and tracked throughout the course of pregnancy to ensure healthy gestational development and assess potential risks.[25]

BMI[V445]: - Body Mass Index is widely recognized as a more effective indicator for assessing whether someone is overweight or underweight. (Bhattacharya, 2007). BMI was categorized as underweight (BMI $\leq$ 18.5), normal (BMI = 18.6–25.0), overweight (BMI = 25.1–30.0), and obese (BMI $\geq$ 30.1) (Doherty DA, 2006).

Anemia (Hemoglobin Level) [V457]: - Described as a reduction in the concentration of hemoglobin in the bloodstream, which leads to significant adverse health outcomes. (Kavsaotlu AR, 2015). Anemia is a global public health concern that affects women of reproductive age, impacting nations regardless of their economic status, spanning across the globe (Habyarimana F, 2020).

ANC Visit during pregnancy[M14$1]: - Described as the initial visit of expectant mothers to prenatal clinics for healthcare services provided by medical professionals. This visit serves the purpose of assessing the well-being of both the mothers and the developing fetus, determining

[23] *https://www.cdc.gov/reproductivehealth/maternalinfanthealth/pregnancy-complications.html*
[24] *https://www.stanfordchildrens.org/en/topic/default?id=newborn-measurements-90-P02673*
[25] *https://www.cdc.gov/reproductivehealth/maternalinfanthealth/pregnancy-weight-gain.htm*

the pregnancy's stage and anticipated delivery date, and laying the groundwork for future check-up schedules. (Gebresilassie, 2019).

STD[V763A]: - Sexually transmitted infections have been linked to a wide range of consequences, which encompass issues such as miscarriage, fetal death, premature birth, low birth weight, post-partum endometritis (Mullick, 2005).

Domestic Violence[D106]: - described as a set of actions within intimate relationship intended to gain or maintain control and power over one's partner. Domestic violence during pregnancy has been linked to a 37% higher likelihood of obstetric problems necessitating prenatal hospitalization (Kaye DK, 2006).

Maternal Age [V447A]: -refers to age of the pregnant women at the time of conception or pregnancy. Pregnancies that occur during adolescence carry elevated risks for both teenage mothers and their infants, primarily due to the young age at which childbirth occurs (Monroy De Velasco, 1982) . A mother is thought to be more likely to experience severe negative pregnancy outcomes if she is considered to have an advanced maternal age, which is typically defined as 35 years or older. (Mehari, 2020.)

Cigarette (Tobacco use) [V463A]: - In particular, regular exposure to tobacco shows a connection with a dose-dependent pattern linked to negative consequences like preterm birth and Low birth weight (Kataoka MC, 2018)

Gestational Age[V233]: - refers to the duration of pregnancy advancement, typically from the first day of the woman's last period to the present day, it is counted in weeks.[26]. It is used to track the development and milestones of the fetus and is essential for monitoring the progression of pregnancy, determining due dates, and assessing the growth and well-being of the fetus. (Unger, 2019).

Place of Delivery[M15$1]: - Place of delivery are independent determinant of adverse pregnancy outcomes. Home delivery increased the risk of adverse pregnancy outcomes by fourfold ((McClure EM, 2015)

[26] *https://www.verywellfamily.com/gestational-age-2371620#toc-gestational-age-at-birth*

Height[V438]: - maternal age has a great impact on the neonate weight and length**.** Short mothers (height ≤145 cm) had 2.74 and 9.0-fold greater risk of delivering LBW baby than average (146–155cm) and tall mothers(>155cm) (Li WU, 2020).

4.3 Data Preprocessing

The gathered datasets consist of a total of 10,220 instances with 15 features. As all these features are the most relevant ones for developing a predictive model that can predict the presence of adverse pregnancy outcome among pregnant women in the case of Ethiopia. Various data preprocessing methods, including tasks like data cleansing, data transformation, addressing class imbalance, eliminating nearly constant features, and employing feature selection and engineering techniques, were utilized. The most common step in data preprocessing is conversion of the original dataset in to CSV file format by using appropriate tools and programming libraries. The next step includes exploratory data analysis, missing value imputation, removal of quasi features, feature selection and scaling. The dataset obtained from Ethiopian Demographic Health Survey contains missing values, imbalanced datasets and categorial values. It necessitates to handle the identified limitations, to improve the quality of a dataset. By addressing these issues, the dataset can be enhanced in terms of its reliability and usefulness.

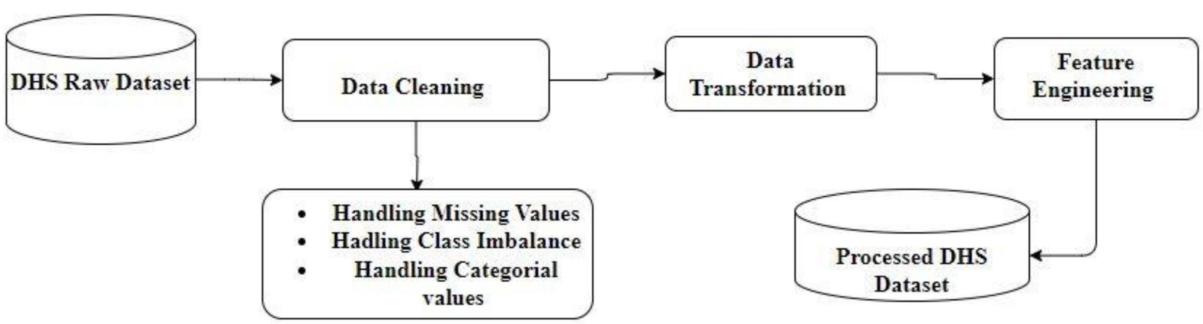

Figure 4-2 Data Preprocessing steps

4.3.1 Handling Missing Values

The missing values were handled using missing value imputation techniques for categorical data. Redundant data was manually eliminated. While quasi-constant features weren't directly removed, we merged them into a single feature. To facilitate the handling of missing values, certain features with multiple distinct values, such as body mass index, wealth index, and height, were converted into discrete values through the process of binning discretization. This transformation allows for easier analysis and interpretation of the data.

Table 4-1 Missing Value Imputation

No.	***Attribute***	***Data types***	***Missing value (%)***	***Remark***
1.	V228	Nominal	0.019%	Mode
2.	M18$1	Nominal	13.874%	Mode
3.	M45$1	Nominal	28.072%	Mode
4.	V438	Numeric	12.29%	Mean
5.	V457	Nominal	15.598%	Mode
6.	M14$1	Nominal	14.7%	Median
7.	V445	Numeric	10.23%	Mean
8.	V763A	Nominal	27.372%	Mode
9.	D106	Nominal	21.436%	Mode
10.	V447A	Numeric	25.942%	Mean
11.	V463A	Nominal	18.31%	Mode
12.	V233	Numeric	0.68%	Median
13.	M15$1	Nominal	2.3%	Mode

4.3.2 Handling Categorical Data

Converting categorical data into numeric form is a crucial step in deep learning for making it compatible with various algorithms that require numeric input. There are several methods for converting categorical data into numeric form. In this approach, every distinct category within a categorical variable is assigned a corresponding numeric identifier. However, it's essential to emphasize that label encoding introduces an arbitrary ordering among the categories, which may not always be meaningful. These are the categorial datasets

Table 4-2 Handling Categorical Data

Attribute name	***Data Type***
Pregnancy Complication	Nominal
Size of the child	Nominal
Iron Supplement	Nominal
Maternal Weight	Numeric
Maternal Height	Numeric
Anemia Level	Nominal
Number Of ANC Visit	Nominal
BMI	Numeric
Any STD	Nominal
Physical Violence	Nominal
Maternal age	Numeric
Cigarette	Nominal
Gestational Age	Numeric
Place of delivery	Nominal

4.4 Feature Selection

Feature selection seeks to improve the performance of a deep learning model by reducing the dataset's complexity, eliminating irrelevant, duplicated, or noisy attributes that could potentially harm the model's accuracy, interpretability, or ability to generalize effectively. These two types of feature selection techniques are used in this study to compare the effectiveness of the features with each other. From these two methods Backward feature selection and Univariate feature selection techniques have performed better than the others.

4.4.1 Backward and Univariate Feature Selection

Feature selection methods that are frequently used in deep learning are backward feature selection and univariate feature selection. It's crucial to keep in mind, though, that deep learning models frequently rely on automatically learning features through neural networks and might not need explicit feature selection. These methods can still be used in the context of deep learning. The backward feature selection and univariate feature selection methods were employed to evaluate various subsets of features from the original feature set. The goal was to assign scores to different feature combinations and identify the subset that achieved the highest score. This process involved iteratively removing features (backward feature selection) or evaluating features individually (univariate feature selection) to determine their effect on the overall performance or relevance of the model. The outcome was the selection of the feature subset with the best score, indicating its potential to contribute effectively to the model's performance or capture relevant information

4.5 Deep Learning Model Building

This objective of the study is to build a deep learning model that predicts adverse pregnancy outcome presence by classifying whether its low birth weight, preterm birth or intrauterine birth based on the features that are collected and selected from EDHS maternal Datasets using deep learning approaches. To address our objective, we build LSTM, FNN, BiLSTM, FNN_LSTM and FNN_BiLSTM based models which are the most recent and repetitively used models' internal health domain. The predictive deep learning model will be developed using FNN with BiLSTM combined. Furthermore, the model's performance is assessed using various evaluation metrics such as accuracy, precision, recall, F1-score, and the confusion matrix.

4.5.1 Hyperparameter Tuning

Grid search is a conventional technique utilized to optimize hyperparameters, involving a systematic exploration on a predefined subset of the algorithm's hyperparameter space. It systematically generates candidate hyperparameter combinations by exhaustively exploring a grid of parameter values. However, grid search can encounter challenges when dealing with high-dimensional spaces, as the number of possible combinations grows exponentially with the number of hyperparameters. Grid search was used to fine-tune the hyperparameters of each algorithm since the algorithm's effectiveness relies heavily on choosing the right hyperparameters. This step has consistently played a pivotal role in the development of deep learning models. (Ramadhan, 2017).

CHAPTER FIVE

5. Implementation and Experimentation

This chapter discusses in detail how the model is implemented and evaluated by using deep learning techniques. The development tools used to predict Adverse pregnancy outcomes are stated in detail with their experimentational results that will address the proposed solution for adverse pregnancy outcome. To build our model, the data collected from EDHS that are preprocessed properly will be used.

5.1 Development Tools

Table 5-1 Development Tools Used

Tools and Packages	Versions	Description
Anaconda Navigator[27]	23.1.0	a desktop graphical user interface that lets you launch applications and manage conda-related tasks without needing to use the command line.
Python [28]	3.9.13	user-friendly platform for learning syntax prioritizes readability, which ultimately lowers the expenses associated with maintaining programs.
Jupiter Notebook[29]	6.4.12	an open-source web app for creating and sharing documents with live code, equations, visualizations, and text.
Pandas[30]	1.4.4	a robust, speedy, and user-friendly open-source data analysis and manipulation tool, built on Python.

[27] *https://docs.anaconda.com/*
[28] *https://www.python.org/doc/essays/*
[29] *https://realpython.com/jupyter-notebook-introduction/*
[30] *https://pandas.pydata.org/*

Matplotlib[31]	3.5.2	versatile Python library that offers extensive capabilities for producing static, animated, and interactive visualizations.
Scikit-learn[32]	1.2.2	freely accessible web-based platform that allows you to create and share documents that combine live code, mathematical equations, graphical representations, and written text.
NumPy[33]	1.23.5	technique employed to transform textual information into numerical format for the purposes of feature extraction, model training, and model evaluation.
Draw.io	21.5.2	Online tool for drawing diagrams, architectures, UML, and other types of drawing
Seaborn[34]	0.11.2	Data visualization library based on matplotlib
Microsoft Excell	2305-2016	For Documentation
Microsoft Word	2305-2016	For Dataset Preparation
TensorFlow[35]	2.12.0	foundational framework for differentiable programming that manages tensors, variables, and gradients.
Keras[36]	2.12.0	graphical interface designed for deep learning that handles elements like models, optimizers, loss functions, metrics, and additional components.

[31] *https://matplotlib.org/*

[32] *https://scikit-learn.org/stable/getting_started.html*

[33] *https://numpy.org/doc/stable/user/whatisnumpy.html*

[34] *https://seaborn.pydata.org/*

[35] *https://www.infoworld.com/ what-is-TensorFlow-the-machine-learning-library*

[36] *https://keras.io/getting_started/intro_to_keras_for_researchers/*

5.2 Experiment Environment

This study used a personal computer equipped with high-performance hardware as its experimental platform. An Intel(R) Core (TM) i7-7500U CPU running at 2.70GHz base clock speed and maximum turbo boost speed at 2.90 GHz CPU served as the device's power source. With 8 GB RAM, the machine has more than enough memory to execute complex computations on large datasets and ensure effective experiment execution. The storage capacity was a 1TB Hard disk drive, enabling quick data access and quicker loading times. The system also has an NVIDIA GeForce GTX 1080 graphics card, which accelerates deep learning model training and evaluation. The entire experimental configuration was supported by the 64-bit Microsoft Windows 10 Pro operating system.

5.3 Dataset Description

The dataset used for adverse pregnancy outcome Prediction using a deep learning approach is derived from the Ethiopian Demographic and Health Survey from the year 2000 till 2018 of the interim EDHS dataset. It has 15 major features and 10220 instances spanning many aspects of pregnancies, such as Gestational age, weight gain during pregnancy, Lifestyle and Environmental style, socioeconomic status, access to prenatal care, maternal health issues, and more, with the key target variable reflecting the outcome of pregnancies. The target variable is a multiclass label that categorizes the outcomes into three categories: "Low Birth Weight", 'Preterm Birth" and "Normal Birth."

The presence of missing values and class imbalance in the dataset's make it difficult to build a reliable and accurate predictive model. By using Imputation or the exclusion of the affected instances both of the problems are dealt through a proper step of preprocessing techniques. Furthermore, the ultimate dataset, which is categorized into three classes, comprises 10,220 rows and consists of 15 columns, including a single column for labeling the target class. as Low Birth Weight, Preterm Birth and Normal Birth.

By utilizing this dataset, the objective of this study is to build a deep learning-based model that can accurately predict adverse pregnancy outcomes based on the features provided. The model has the potential to have a significant impact because timely detection of pregnancies at risk for low birth weight or preterm birth would enable healthcare professionals to act pro-actively and provide targeted care, potentially reducing adverse pregnancy outcomes and improving maternal and neonatal health outcomes in Ethiopia.

Table 5-2 Dataset Description

No.	Symbols	Feature name	Data Type	Description
1.	V228	Pregnancy Complication	Nominal	Any disease or condition that affects the expectant mother's pregnancy.
2.	M18$1	Size of the child	Nominal	weight at birth of less than 2.5 kg
3.	M45$1	Iron Supplement	Nominal	to prevent APO, it is recommended that women in their childbearing years should aim for a daily intake of 400 micrograms of folic acid.
4.	V437	Maternal Weight	Numeric	the weight of a woman's body while she is expecting a child
5.	V438	Maternal Height	Numeric	has great impact on the neonate weight and length
6.	V457A	Anemia Level	Nominal	reduction in the amount of red blood cells or the amount of hemoglobin in the bloodstream. Women's that have iron deficiency anemia are at great risk of having an adverse pregnancy.
7.	M14$1	Number Of ANC Visit	Nominal	Supervision of humans during pregnancy to monitor the progress of fetal growth
8.	V445	BMI	Numeric	During pregnancy having underweight, overweight or obese BMI values can cause a severe adverse pregnancy outcome for both the mother and the unborn child.

9.	V763A	Any STD	Nominal	pose challenges during pregnancy and potentially result in significant repercussions for both the pregnant woman and the developing fetus.
10.	D106	Physical Violence	Nominal	a recurring set of actions within a relationship aimed at acquiring or retaining dominance and authority over a romantic partner.
11.	V447A	Maternal Age	Numeric	refers to age of the pregnant women at the time of conception or pregnancy.
12.	V463A	Cigarette	Nominal	Smoking while pregnant increases the risk of premature birth, a significant contributor to infant mortality, disability, and health issues.
13.	V233	Gestational Age	Numeric	describes how far the pregnancy is.
14.	M15$1	Place of delivery	Nominal	Promoting institutional childbirth is advocated to reduce the chances of unfavorable pregnancy outcomes.

5.4 Preprocessing and Model Implementation

The presence of missing values and imbalanced class has impacted by the performance of the model. Whenever we design a deep learning model, preprocessing approaches are the major steps that will be utilized in our study to ensure the quality of the dataset. Furthermore, the dataset we get from EDHS were not a complete, clean and well organized. Therefore, by exploring the datasets we were able to utilize missing values and class imbalances by using various preprocessing approaches like handling missing values, handling categorical values, removing redundant data and the likes.

5.4.1 Importing Libraries

Before starting the preprocessing, we need libraries and packages that needs to be imported to implement the proposed solution. These libraries are a collection of pre-built python modules that are used for a various function. In our case these are the libraries we have used to implement the proposed model.

```
import pandas as pd
import tensorflow as tf
import matplotlib.pyplot as plt
import seaborn as sns
import mlxtend
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler
from tensorflow.keras.models import Sequential
from sklearn.metrics import accuracy_score,roc_curve, auc
from tensorflow.keras.optimizers import Adam, RMSprop
from sklearn.metrics import precision_score, f1_score, confusion_matrix,classification_
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, LSTM, Bidirectional, Flatten, SimpleRNN
from sklearn.model_selection import StratifiedKFold , cross_val_score
from tensorflow.keras import regularizers
from keras.utils import to_categorical
from sklearn.feature_selection import SelectKBest,f_classif, RFECV,mutual_info_classif,
from sklearn.neural_network import MLPClassifier
from sklearn.feature_selection import SelectFromModel
```

Figure 5-1 Importing Necessary Libraries and Packages

5.4.2 Load the Dataset

Once the libraries and packages are imported, the next step is loading the dataset we have collected from EDHS by using pandas library for further analysis and exploration. Our dataset have a CSV format that will make it more suitable to be loaded to the dataframe without any further steps. The code below shows the loading of our dataset.

```
#import neccessary packages  and load the dataset
import pandas as pd
# Load the  dataset
dataset = pd.read_csv('C:/Users/Fozi/Desktop/BCD/RefinedBook1.csv')
```

Figure 5-2 Implementation of Loading Unrepressed Dataset Using Pandas

5.4.3 Handling Missing Values

By using missing value handling techniques, we can ensure the effective Prediction of the model. Mode imputation techniques were used to address missing values in categorical data. To fill the missing values of the dataset with missing value imputation techniques we have used fillna () method from pandas' library that is utilized to complete missing values in the data frame. It allows us to replace the missing values by using mean and median imputation techniques.

```
#import the neccessary packages for handling missing values
import pandas as pd
from sklearn.impute import SimpleImputer
dataset = pd.read_csv('C:/Users/Fozi/Desktop/BCD/RefinedBook1.csv')
missing_values = dataset.isnull().sum()
print(missing_values)
imputer = SimpleImputer(strategy='mean', missing_values=np.nan) #'median','most_frequen
imputer.fit(dataset)
dataset = imputer.transform(dataset)
```

Figure 5-3 Implementation of Handling Missing Values Using Imputation Approaches

5.4.4 Handling Categorical Values

Converting nominal data into numeric is one of the steps that are addressed during the preprocessing phase of the dataset. By assigning every unique category of the categorial values in to numeric labels. Therefore, the categorical values are managed and put in to practice as explained in detail in chapter 4

```
from sklearn.preprocessing import OneHotEncoder
import pandas as pd
dataset= pd.read_csv('C:/Users/Fozi/Desktop/BCD/RefinedBook1.csv')
categorical_columns=['V228','M14$1', 'M15$1', 'M18$1', 'M45$1', 'V457','V463A','V763A'
                     ,'D106']
encoder = OneHotEncoder (sparse=False)
encoded_feature= encoder.fit_transform(dataset[categorical_columns])
```

Figure 5-4 Handling Categorical values

In the above code we have imported the necessary libraries packages. The categorical columns contain the list columns that contains categorical features from the datasets. By creating the instance of OneHotEncoder that has a sparse parameter that indicates the encoded features will be returned as dense matrix. The next line uses an encoder.fit_transfrom () that fits the encoder on the selected categorical columns of the dataset and transform them in to encoded features. After the dataset is well organized and preprocessed, the features will split in to dependent and independent variables. The dataset without feature selection will be used as X and the dataset that contains the label or the class will be used as Y. For training the model we have used X and Y for the entire dataset. By using fit () method for each model classes we have trained the dataset on various deep learning approaches.

```
#import the neccessary libraries and packages
import pandas as pd
# Load the preprocessed dataset
dataset = pd.read_csv('C:/Users/Fozi/Desktop/BCD/RefinedBook1.csv')
# Separate the independent variables (X) and dependent variable (y)
X = dataset.drop('outcome', axis=1)#represents independent variable
y = dataset['outcome']  #represnets dependent variable
```

Figure 5-5 Splitting Independent and Dependent Features

5.4.5 Feature Scaling

A preprocessing step in deep learning that ensures all the features has similar scales. This will enhance the model's performance and convergence, particularly for models that are influenced by the scaling of input data.

```
# Scale the input data
scaler = StandardScaler()
X = scaler.fit_transform(X)
```

Figure 5-6 Scaling Dataset

In the above code X represents the data we wanted to scale. The initial line is utilized to instantiate the Standard Scaler class that standardizes features by subtracting the mean and scaling to unit variance. The fit transform () method performs two operations. First it fits the scaler to the data and transform the data by applying the scaling. The method in the second line calculates the mean and standard deviation of each feature in X and applies the standardization transformation to the data. The transformed data are assigned to the scaled variable**.**

5.4.6 Label Encoding

By utilizing Label Encoding, categorical columns are converted into numeric format, enabling compatibility with deep learning models that exclusively process numerical data. It is an important pre-processing step in a deep-learning project[37]. For our target variable that has a categorical values, label encoder is used to replace the categorical values to the encoded ones.

```
import pandas as pd
from sklearn.preprocessing import LabelEncoder
# Load and preprocess the dataset
dataset = pd.read_csv('C:/Users/Fozi/Desktop/BCD/RefinedBook1.csv')
# Extract the target variable
target = dataset['outcome']
# Create an instance of LabelEncoder
label_encoder = LabelEncoder()
# Fit and transform the target variable
encoded_target = label_encoder.fit_transform(target)
```

Figure 5-7 Label Encoding

In the code the target variable is extracted from the dataset. Instance of the label encoder is created following label_encoder.fit_transform method () which is responsible to fit the label to the target variable and it will transform it to the encoded values. Moreover, the original target variable will be replaced by the encoded values.

5.5 Deep Learning Model Implementation

The preprocessed dataset will now be used to train the model. For this study we have used various deep learning approaches with different hyperparameter tunning methods. Further implementation of possible architectures proposed in this study are mentioned below. Before creating and building the model for our dataset, the independent and dependent variable as X and Y respectively needs to be split. X contains all the features of the dataset which is encoded and Y is the labeled target class variable.

After the dataset is preprocessed and splits into independent and dependent features, the next step is building deep learning algorithms and training the model with the training dataset that

[37] *https://www.geeksforgeeks.org/ml-label-encoding-of-datasets-in-python/*

will fulfill the objective of this study. The proposed deep learning model FNN_BiLSTM is compared with other deep learning approaches. For this study we have used 5 experiments including LSTM, BiLSTM, FNN, FNN_LSTM and FNN_BiLSTM. The proposed model will be evaluated using evaluation metrics along with stratified 10 k fold cross validation. The necessary packages and libraries used for the proposed model are as follow

```
# Split the dataset into features and target variables
X = dataset.drop('outcome', axis=1)  # Independent variables
y = dataset['outcome']  # Dependent variable
```

Figure 5-8 Splitting the Dataset

The independent variables are extracted in to the 'X' data frame by dropping the outcome column by using drop () method. The dependent variable is assigned to 'Y' series by selecting the 'outcome' column.

5.5.1 Create a Model

Keras is used to build a model. By importing Keras that uses TensorFlow as the back end. We utilized the sequential () module from Keras to initialize the artificial neural network, which is a computational network constructed with inspiration drawn from the human brain. To add a layer to our deep learning model we have imported dense module. When we build a deep learning model, we are usually specifying three layers.[38]

The input layer is the layer through which we will input the features from our dataset. It serves as a passing path for the features to the hidden layer without any computation occurrence. The hidden layer are the layers between the input layer and the hidden player. Hidden layers can be one or more depending on the specific type of problem we have. These layers conduct computations and transmit information to the output layer. The output layer's role is to depict the neural network that will yield the desired results once our model is trained. These layers are in charge of generating the output variables.

[38] *https://www.digitalocean.com/how-to-build-a-deep-learning-model-to-predict-employee-retention-using-keras-and-tensorflow*

```
model = Sequential()
```

Figure 5-9 Creation of a Model

In the above code we have used a sequential () class of TensorFlow Keras API to initialize a linear stack of layers to build deep learning models. Since our problem is a Prediction, we can be able to make predictions from labeled datasets. After initializing the model, we can add layers to the models by using ‘add ()’ method.

```
# LSTM model with 2 hidden layers
lstm_model = Sequential()
lstm_model.add(LSTM(32, input_shape=(14, 1), activation='relu'))
lstm_model.add(Dropout(0.2))
lstm_model.add(Dense(32, activation='relu'))
lstm_model.add(Dropout(0.2))
lstm_model.add(Dense(3, activation='softmax'))
```

Figure 5-10 Implementation of LSTM

In the above code, the first line initializes the LSTM model. For this model we have used two hidden layers by using add () method. In the second line the initial parameter represents the number of nodes of networks 32 is used. The second parameter states the number of features in our datasets. The third parameter is an activation function by which our deep learning model learns by using ReLU. The third line states that we have used a drop out regularization to overcome overfitting by adding a dropout rate between input and output layers. The fourth and fifth lines are the additional layers that are added to the hidden layers. The sixth line states the number of output nodes followed by the activation function of SoftMax.

```
# Feed Forward Neural Network (FNN) model
fnn_model = Sequential()
fnn_model.add(Dense(32,input_shape=(14, 1),activation='relu',
                    kernel_regularizer=regularizers.l2(0.01)))
fnn_model.add(Flatten())
fnn_model.add(Dense(32, activation='relu', kernel_regularizer=regularizers.l2(0.01)))
fnn_model.add(Dense(3, activation='softmax', kernel_regularizer=regularizers.l2(0.01)))
```

Figure 5-11 Implementation of FNN

In the above code, the first line initializes the FNN model. In the second line the first layer is a dense layer with 32 units of neurons which have a total number of 14 input features and 1 target value for each feature. In the same line we have used an activation function of ReLU followed with a kernel regularize that is used for regularizing the weight of the layer with a rate of 0.01. The third line added a flatten layers that flattens the input. The next line added a dense layer with 32 units of neurons with ReLU activation and weight regularization rate. The last line is the output layer with dense layer having 3 of the multiclassification of our data units with activation function of SoftMax that is responsible for the distribution of the 3 classes with weight regularize rate of 0.01.

```
# BiLSTM model
bilstm_model = Sequential()
bilstm_model.add(Bidirectional(LSTM(128, input_shape=(14, 1), activation='relu',
                                    kernel_regularizer=regularizers.l2(0.01))))
bilstm_model.add(Dropout(0.3))
bilstm_model.add(Dense(128, activation='relu'))
bilstm_model.add(Dropout(0.3))
bilstm_model.add(Dense(3, activation='softmax',
                       kernel_regularizer=regularizers.l2(0.01)))
```

Figure 5-12 Implementation of BiLSTM

Certainly, the selection of hyperparameter values in deep learning models is a crucial and often intricate process that requires a blend of experience, experimentation, and adherence to best practices. In the provided code, several hyperparameter choices are made, each serving a specificpurpose. For instance, the decision to employ 128 neurons in the LSTM layer could have been influenced by the model's intended capacity and the dataset's complexity. The use of the RectifiedLinear Unit (ReLU) activation function in hidden layers is a common practice due to its ability tointroduce non-linearity and capture complex patterns. Weight regularization with a rate of 0.01 isintroduced to curb overfitting by penalizing large weights. Similarly, the dropout rate of 0.3 helps prevent overfitting by randomly deactivating a portion of neurons during training. These choices align with well-established practices in deep learning but are not set in stone. Fine-tuning these hyperparameter values may be necessary, as their impact can vary based on the specificcharacteristics of the dataset and the nuances of the problem at hand.

```python
# FNN with LSTM model
fnn_lstm_model = Sequential()
fnn_lstm_model.add(LSTM(64, input_shape=(14, 1), activation='relu',
                        kernel_regularizer=regularizers.l2(0.04)))
fnn_lstm_model.add(Dropout(0.6))
fnn_lstm_model.add(Dense(64, activation='relu',
                        kernel_regularizer=regularizers.l2(0.04)))
fnn_lstm_model.add(Dropout(0.6))
fnn_lstm_model.add(Dense(3, activation='softmax',
                        kernel_regularizer=regularizers.l2(0.04)))
```

Figure 5-13 Implementation of FNN_LSTM

The model is created by using sequential () class from Keras. The second line adds a dense layer with 64 units of neuron numbers that indicates the input feature having a shape of 14 features each having a target value followed by an activation function of ReLU and a weight regularization strength for the FNN_LSTM layer with a rate of 0.04. The third line adds a dropout rate of 0.6 that indicates that 60% of the neurons will be deactivated. Finally, the output layer adds a dense layer with 3 number of neurons indicating a Prediction problem followed by SoftMax activation function and a weight regularization strength for the FNN_LSTM layer with a rate of 0.04.

```python
# FNN with BiLSTM model
fnn_bilstm_model = Sequential()
fnn_bilstm_model.add(Bidirectional(LSTM(128, input_shape=(14, 1), activation='relu')))
fnn_bilstm_model.add(Dropout(0.8))  # Add dropout layer
fnn_bilstm_model.add(Dense(128, activation='relu'))
fnn_bilstm_model.add(Dropout(0.8))  # Add dropout layer
fnn_bilstm_model.add(Dense(3, activation='softmax'))
```

Figure 5-14 Implementation of FNN_BiLSTM

The above code the first line initializes the FNN_BiLSTM model. The second line adds a dense layer that wraps a BiLSTM layer with 128 units of neurons followed by the number of input features by using ReLU activation function that will have merge mode that concatenates the output of forward and backward LSTM layer. The third line adds a dropout rate of 0.8 that indicates that 80% of the neurons will be deactivated. The fourth layer adds a dense layer. The final line adds a dense layer that is responsible for the output layer having 3 number of units using SoftMax activation function.

5.5.2 Compile the Model

In Keras, the model offers a compile() method for preparing the model for training.The argument and default values contains loss , optimizer and metrics values. After completing the model creation process, the next step is to compile it. Once the compilation is done, the subsequent step involves training the model[39]. The purpose of the loss function is to identify errors or deviations during the learning process. Keras requires a loss function during its compilation and for our study we have used categorical_crossentropy that is well suited for a muli-class Prediction problem. Optimization is a significant procedure that fine-tunes the input weights by comparing the model's predictions with the loss function. From the various optimization options keras provides, we have used adam optimizer for our study. The metrics is used to evaluate the accuracy of the our model.

```
model.compile(loss='categorical_crossentropy',
              optimizer=keras.optimizers.Adam(learning_rate=0.001), metrics=['accuracy'])
```

Figure 5-15 Model Compilation

5.5.3 Model Training

Models are trained by NumPy arrays by using the fit () methods. The fit () function will evaluate the model over training.

```
history = model.fit(X_train, y_train, validation_data=(X_val, y_val),
                    epochs=50, batch_size=20 , verbose=1)
```

Figure 5-16 Model Training

In this above code the model is trained by using the fit (). The first parameters are a TensorFlow tensor training data inputs and their labels respectively. The second parameters are representing the validation data input and labels respectively. At the end of each epoch the model will evaluate its performance on this data. The third parameter signifies how many times the model goes through the entire training dataset in iterations. The fourth parameter defines the number of samples to be processed in each training dataset. The last parameter

[39] *https://www.digitalocean.com/ how-to-build-a-deep-learning-model*

controls the verbosity of the model. In our case the training run by epoch of 50 and batch size of 20.

5.6 Model Evaluation

In this research, various evaluation metrics were applied to assess the performance of the deep learning model using the preprocessed dataset. Model evaluation can be addressed by using various evaluation metrics. From those we have used Stratified 10 K-fold cross validation to evaluate the performance of our model. Stratified K-fold cross-validation is a technique used for cross-validation that differs from the traditional K-Fold approach. It generates stratified folds, ensuring that each fold maintains the same proportion of samples for every class. [40]. From Prediction evaluation metrics we have used accuracy, precision, f1-score and confusion matrix Additionally, the model's performance is assessed using stratified k-fold cross-validation with a value of k set to 10.

```
kfold = StratifiedKFold(n_splits=10, shuffle=True, random_state=42)
fold_accuracy = [] # Calculate and print the average accuracy for each fold
for i, accuracy in enumerate(fold_accuracy):
    print(f"Fold {i+1} Accuracy: {accuracy*100:.2f}%")
# Calculate and print the average accuracy across all folds
avg_accuracy = sum(fold_accuracy) / n_splits
print(f"Average Accuracy: {avg_accuracy*100:.2f}%")
```

Figure 5-17 Stratified K-Fold Cross Validation

[40] *https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html*

CHAPTER SIX

6. Result, Evaluation and Discussion

This chapter provides an exposition of the experimental outcomes and findings. The overview of the dataset with their respective feature importance and distribution, model performance and evaluation results including activation function and algorithm function are discussed in this section of the study. The study also includes a comparison between previous works and the proposed model for adverse pregnancy outcome using deep learning approaches. The chapter also includes a discussion that focuses on how the proposed study answers the research questions.

6.1 Dataset Distribution and Prediction

After preprocessing the raw data collected from EDHS, the study utilized a total of 10,220 instances with 15 columns, including the target outcome class. The study of Adverse pregnancy outcome Prediction classifies pregnancy outcomes in to three class as Normal Birth, Low Birth weight and Preterm Birth. The figure below indicates the distributions of classes collected from the EDHS dataset after preprocessing with three classes having a class imbalance in the dataset indicating the instances are not equally distributed among the target class. In order to fix class imbalances, the study has used over-sampling, under-sampling, and SMOTE techniques to achieve a better-balanced class distribution for the models. The over-sampling techniques allowed us to focus on the duplication of instances from the minority class to increase the representation in the dataset and under-sampling techniques allowed us to remove instances from the majority class to decrease their representation in the dataset. Furthermore, Synthetic Minority Over-Sampling Technique allowed us to create synthetic samples from the minority classes by using the existing instances. These techniques have been used in the study to balance the class distribution without simply duplicating instances. By using these datasets, the overall results of the proposed models are summarized as follows. The three-class distribution of the dataset is stated in the below figure.

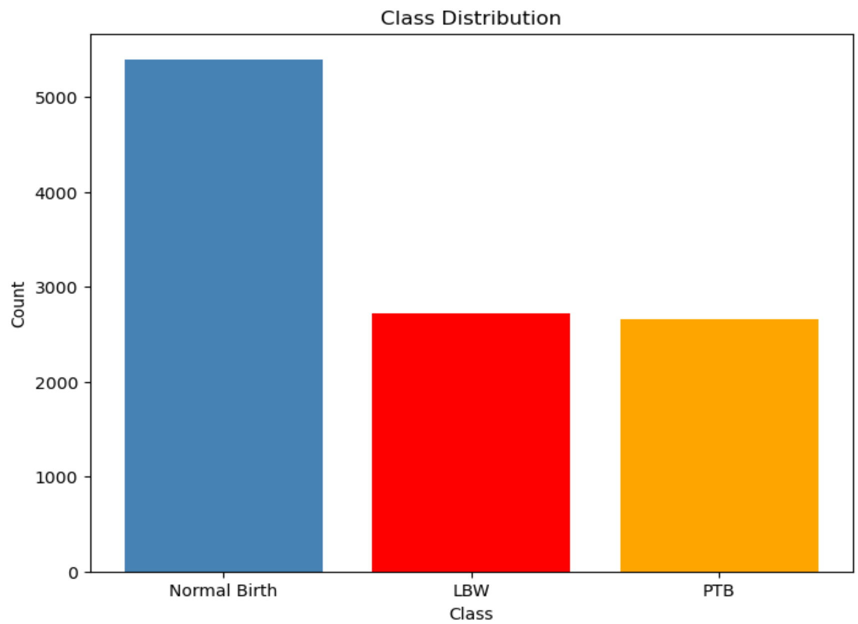

Figure 6-1 Class Distribution of the Dataset

6.2 Feature Importance

The figure below presents the score of features according to their importance. The relevance of the features indicates how important the features are in contribution of the target Prediction for adverse pregnancy outcome Prediction using deep learning approaches. The feature that has highest importance score are considered to be the features that have a biggest impact on the accuracy of the target features. The most important function are the features that have highest importance score compared with the lowest score. In our study pregnancy complication (V228) has the highest importance score and Anemia level (V457) is the least important one.

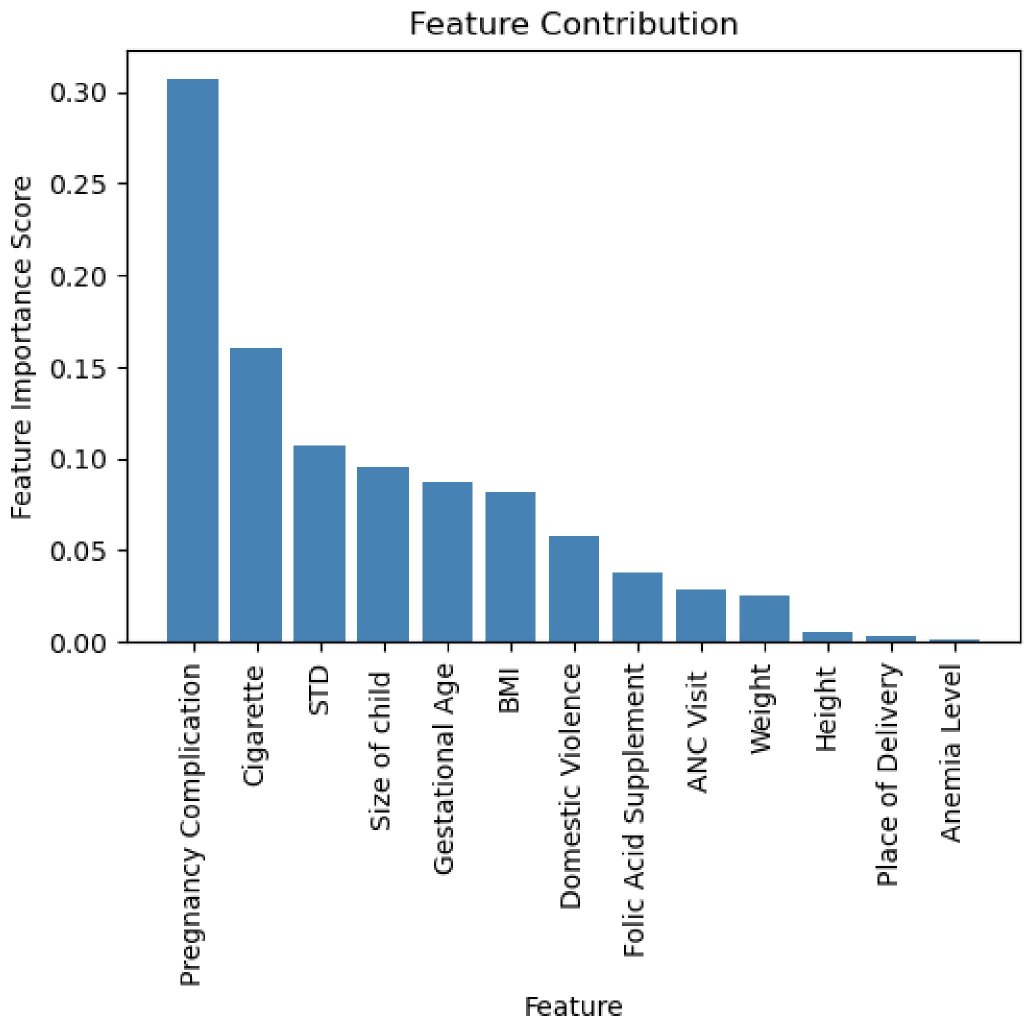

Figure 6-2 Feature Importance Score

6.3 Results and Performance of the Proposed Model

In this study a variant of deep learning approaches has been experimented and presented. From the variant models LSTM, BiLSTM, FNN, FNN_LSTM and FNN_BiLSTM, were experimented by using various hyperparameter tunning. From these variants of RNN based models LSTM, FNN and their hybrid FNN_LSTM has been presented and evaluated to predict adverse pregnancy outcome using EDHS dataset. The summarized result of each of the deep learning models are as follows.

6.4 Performance of the Model

The performance evaluation of the proposed model for FNN, BiLSTM, LSTM, FNN_BiLSTM and FNN_LSTM based on the dataset are discussed below.

6.4.1 Confusion Matrix of FNN Model

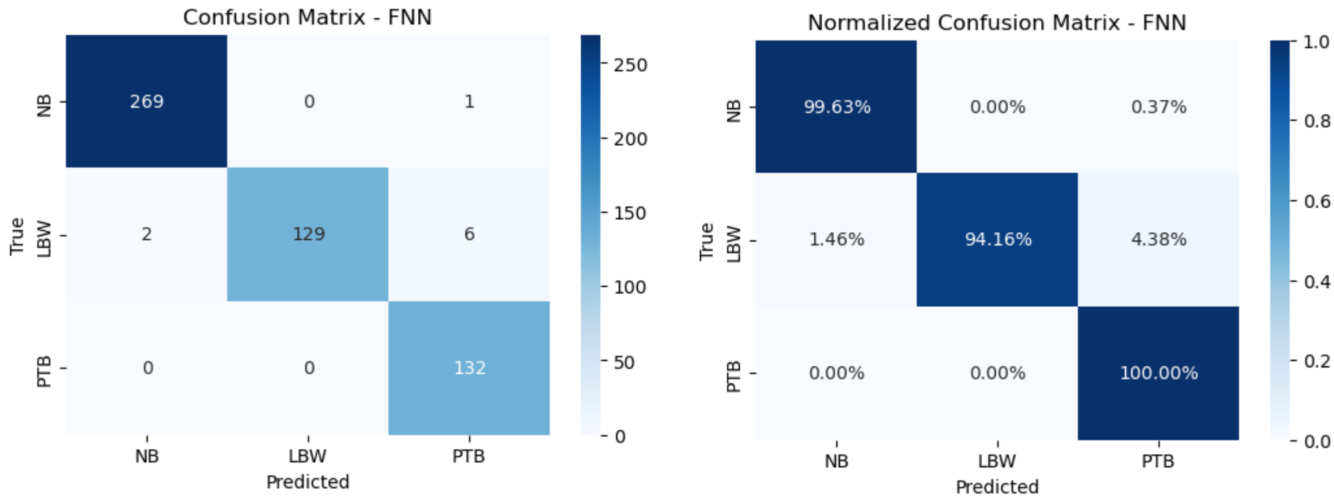

Figure 6-3 Confusion Matrix of FNN Model with their Respective Normalization

In the figure displayed above, the FNN model's performance was evaluated on the EDHS dataset using a stratified 10-fold cross-validation approach. The confusion matrix results indicate that, out of the total instances, 269 were correctly classified as Normal Birth, 129 as LBW, and all 132 instances as PTB. The algorithm misclassified instances of 1 from Normal Birth as PTB, 2, 6 instances as Normal Birth and PTB respectively. The right-side depiction of the normalized confusion matrix presents accurately classified instances in percentage form. This normalized confusion matrix is identical to the one shown on the left side.

6.4.2 Confusion Matrix of BiLSTM model

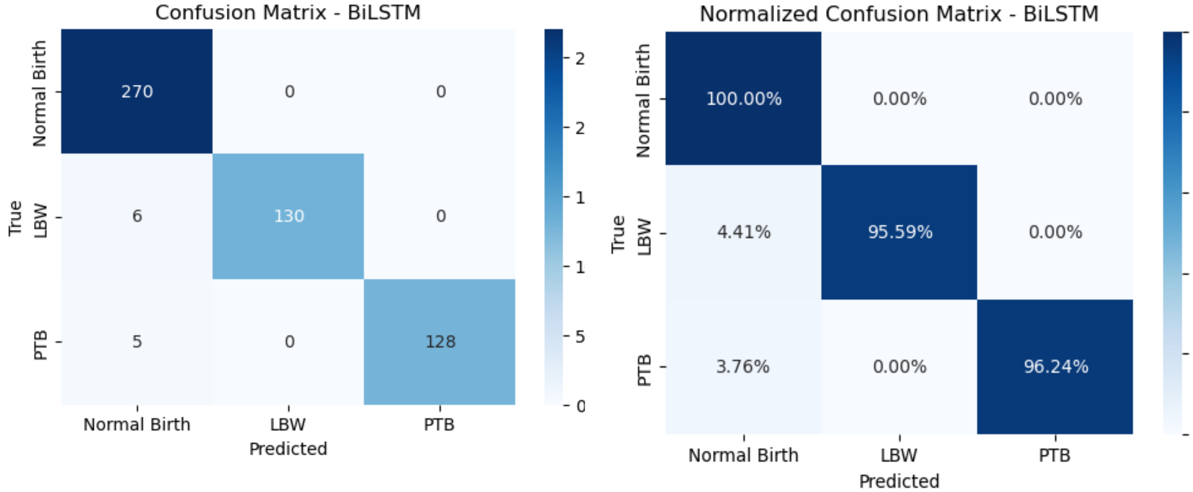

Figure 6-4 Confusion Matrix of BiLSTM model with their Respective Normalization

In the above shown figure the performance of BiLSTM model experimented on EDHS dataset using stratified 10-fold confusion matrix provides as such 270 out of 270, 130 out of 136 and 128 out of 132 dataset instances are correctly classified as Normal Birth, LBW and PTB respectively. The algorithm misclassified instances are 6 from LBW as Normal Birth and 5 instances from PTB as Normal Birth. The right-side depiction of the normalized confusion matrix presents accurately classified instances in percentage form. This normalized confusion matrix is identical to the one shown on the left side.

6.4.3 Confusion Matrix of LSTM model

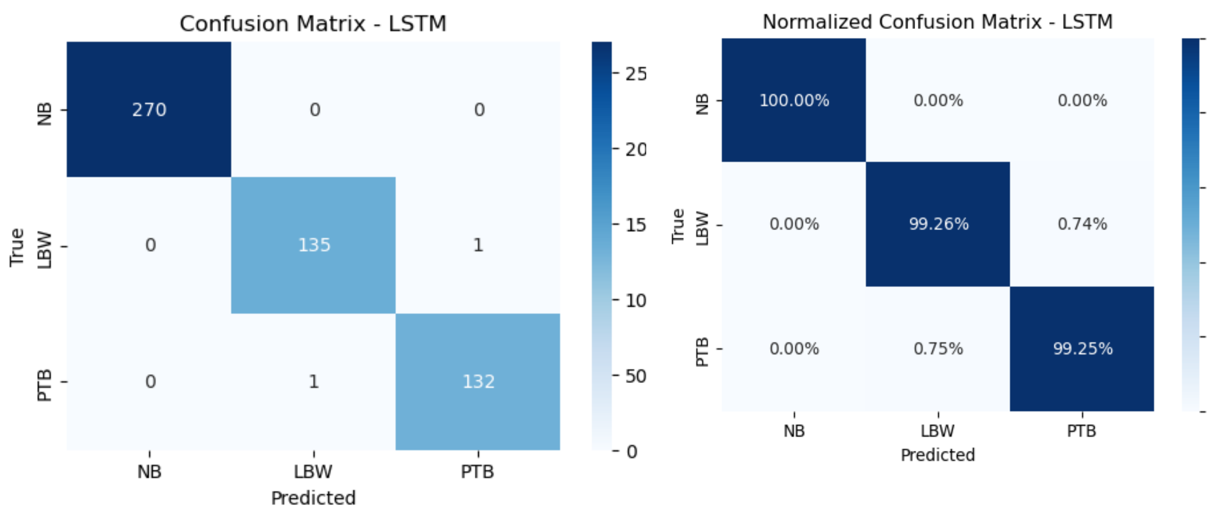

Figure 6-5 Confusion Matrix of LSTM model with their Respective Normalization

In the above shown figure the performance of LSTM model experimented on EDHS dataset using stratified 10-fold confusion matrix provides as such 270 out of 270, 135out of 136 and 132 out of 133 dataset instances are classified as Normal Birth, LBW and PTB respectively. The algorithm misclassified 1 instance from LBW as PTB and 1 instance from PTB as LBW. The right-side depiction of the normalized confusion matrix, indicates classified instances in the form of percentage. It also indicates the percentage of misclassified instances. The normalized confusion matrix is the same with the left side illustration.

6.4.4 Confusion Matrix of FNN_LSTM model

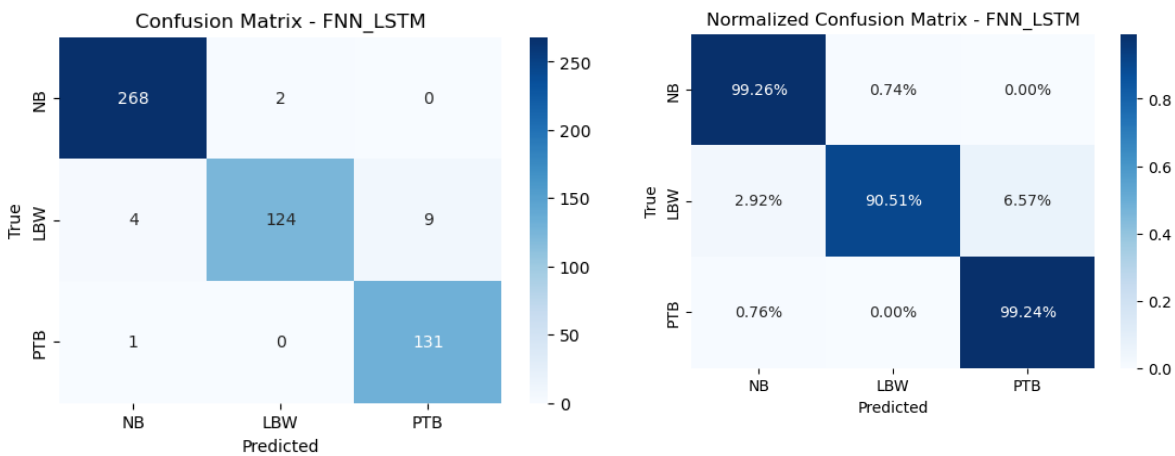

Figure 6-6 Confusion Matrix of FNN_ LSTM model with their Respective Normalization

In the above shown figure the performance of FNN_LSTM model experimented on EDHS dataset using stratified 10-fold confusion matrix provides as such 268 out of 270, 124 out of 137 and 131 out of 132 dataset instances are classified as Normal Birth, LBW and PTB respectively. The algorithm misclassified 2 instances from Normal Birth as LBW, 4 and 9 instances from LBW as Normal Birth, 9 from LBW as PTB respectively. The final class has 1 misclassified from PTB as Normal Birth. The right-side depiction of the normalized confusion matrix presents, classified instances in the form of percentage. It also indicates the percentage of misclassified instances. The normalized confusion matrix is the same with the left side illustration. The only difference between the two figures is the one on the left side is without normalization and the right side one has a normalized confusion matrix.

6.4.5 Confusion Matrix of FNN_BiLSTM model

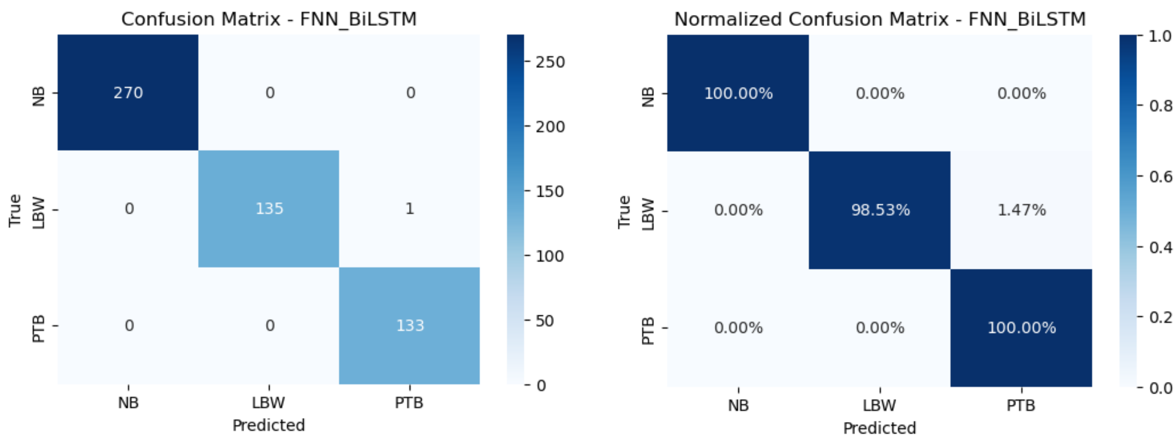

Figure 6-7 Confusion Matrix of FNN_ BiLSTM Model with their Respective Normalization

In the above shown figure the performance of FNN_BiLSTM model experimented on EDHS dataset using stratified 10-fold confusion matrix provides as such 270 out of 270, 135 out of 136 and 133 out of 133 dataset instances are classified as Normal Birth, LBW and PTB respectively. The algorithm misclassified 1 instance from LBW as PTB. The occurrence of misclassification within health-related datasets was attributed to the presence of class overlaps. This phenomenon stems from the inherent variability in health data, where individuals across different health categories may exhibit similarcharacteristics or measurements. Health conditions often manifest in multifaceted ways, with symptoms that can overlap or mimic those of unrelated ailments, making accurate Prediction aformidable task. Adverse pregnancy outcomes are influenced by a multitude of interconnected factors, such as maternal health, genetics, lifestyle, and environmental variables. These factors can lead to instances where pregnancies with similar features or risk factors may still result in differentoutcomes due to subtle underlying factors that are not easily discernible. Consequently,the presence of class overlap occurs when instances from different classes share common features,making it challenging for the model to make accurate predictions. Collaboration with domain experts and healthcare professionals is essential to navigate the intricacies of health data and enhance Prediction accuracy while managing overlaps effectively (Alogogianni, 2023). In a task where we need to classify different types of outcomes if two or more of these types have similar characteristics, it becomes challenging to establish clear distinctions between them using the same set of features. As a result, the classifier's ability to accurately tell apart these overlapping classes diminishes (Wahba, 2020). The right-side depiction of the normalized confusion matrix presents, classified instances in the form of percentage.

6.5 Results of the proposed model

6.5.1 Result of FNN model

The simplest ANN architecture is the feedforward neural network (FNN), which stacks the neurons layer by layer in a feedforward manner. Using FNN proposed model, the study has achieved 95.47% of accuracy under stratified 10-fold cross validation. The figure below illustrates the training, validation and loss curves of the model.

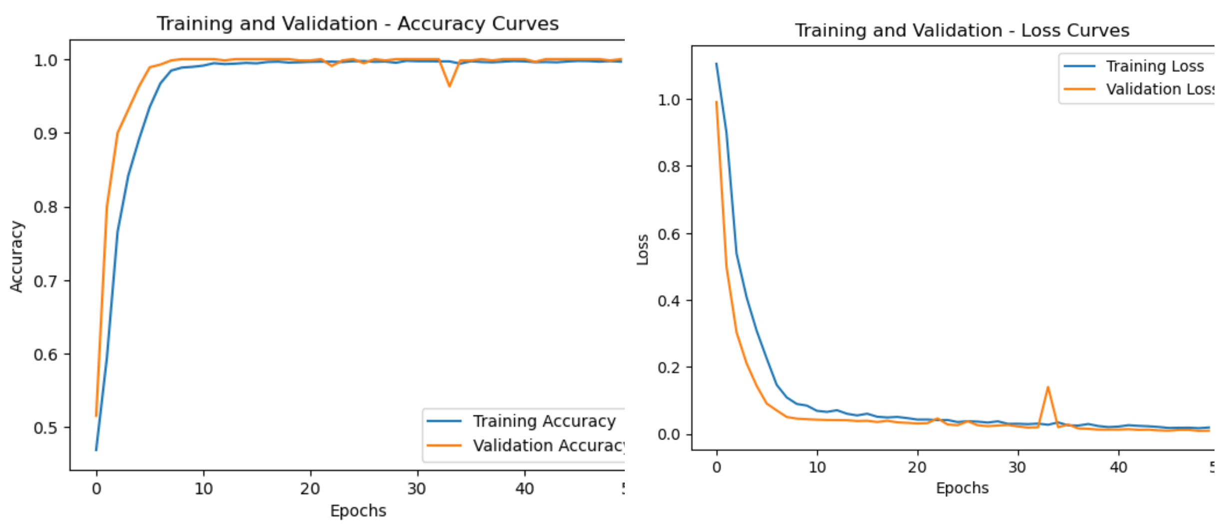

Figure 6-8 Accuracy and Loss Curve of FNN Model Through Epochs

6.5.2 Results of BiLSTM

In bidirectional, the input flows in two directions, making a Bi-LSTM different from the regular LSTM. The output value is computed by using both forward and backward LSTM layers. By using this proposed BiLSTM model the study has achieved 96.99 % of accuracy under stratifies 10-fold cross validation. The figure below presents the training, validation and loss curves of the model.

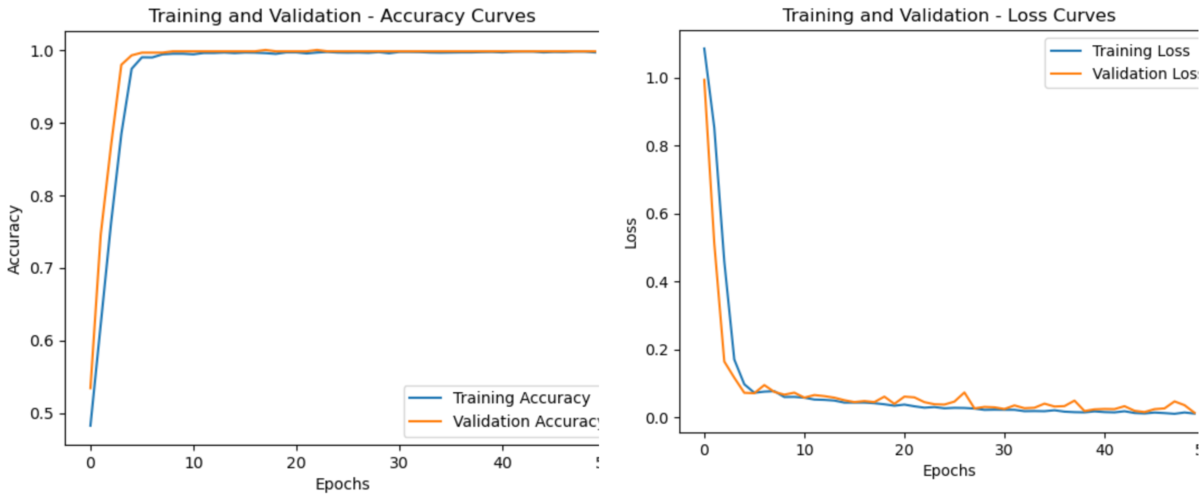

Figure 6-9 Accuracy and Loss Curve of BiLSTM Model Through Epochs

6.5.3 Results of LSTM

A type of RNN model architecture that is widely used for handling sequential data. They are designed to overcome the limitation of RNN by capturing long-term dependencies of data. The study used LSTM model and achieved 97.29% of accuracy under stratified 10-fold cross validation. The figure below illustrates the training, validation and loss curves of the model.

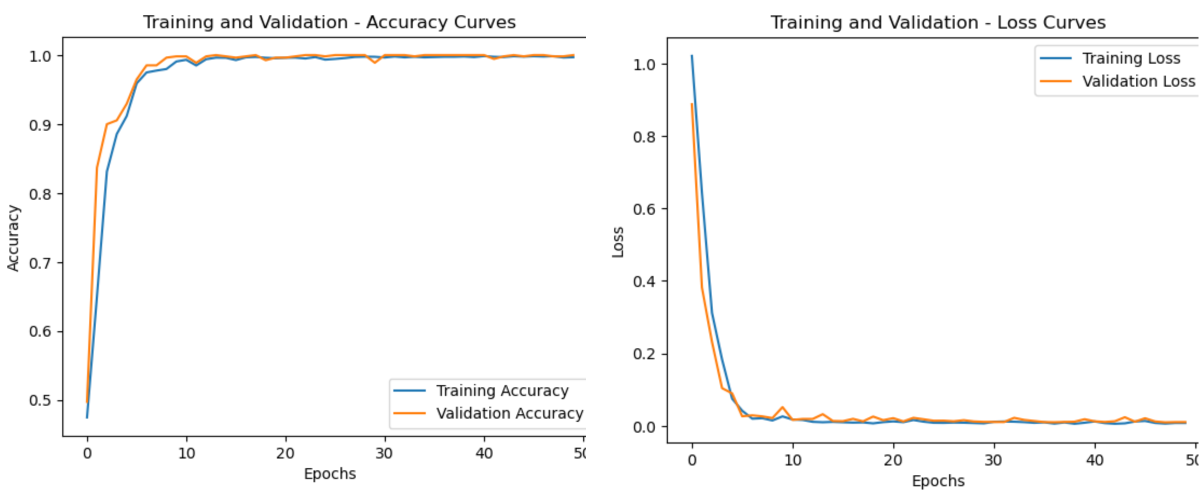

Figure 6-10 Accuracy and Loss Curves of LSTM Model Through Epochs

6.5.4 Results of FNN_BiLSTM

In this study we have used the combination of FNN with BiLSTM. The combination of this models will improve the limitations of both models. This study used a bidirectional wrapper that creates several LSTM layers. These layers process both the input sequence in the forward order and the output in the backward order. The final results will be concatenated by providing access to the past and future context of the dataset. From the proposed deep learning models FNN_BiLSTM outperformed the rest models with an accuracy of 98.94% under stratified 10-fold cross validation. The figure below presents the training, validation and loss curves of the model.

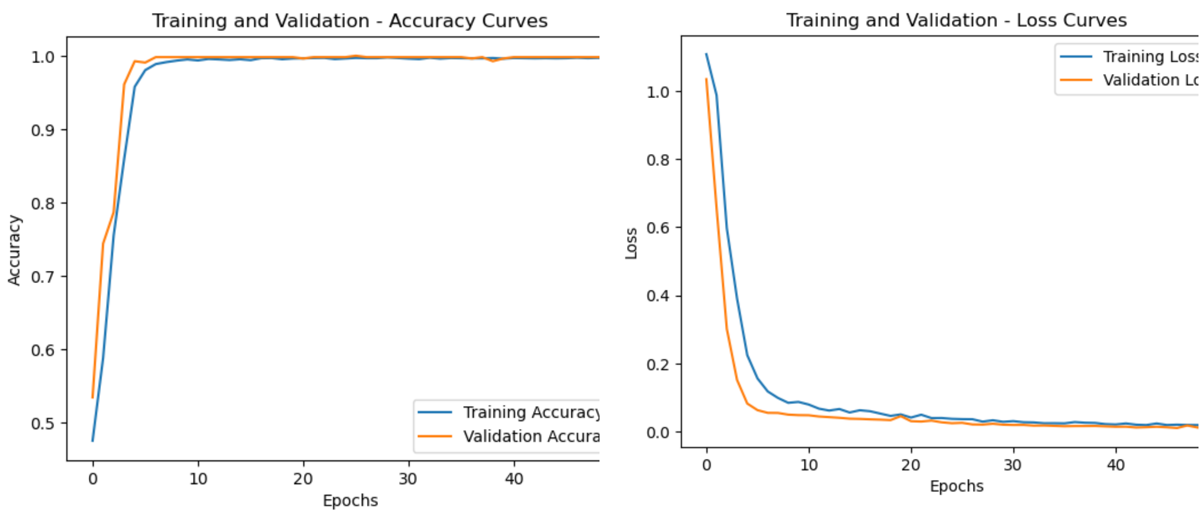

Figure 6-11 Accuracy and Loss Curve of FNN_BiLSTM Model Through Epochs

6.5.5 Results of FNN_LSTM

The hybrid combination FNN with LSTM was used in this study. These models capture the non-linear data in the dataset and while the LSTM layer handles sequential dependencies and captures long-term dependencies of the of the dataset in the input sequence. By using FNN_LSTM the study has achieved an accuracy of 92.49 % under stratified 10-fold cross validation. The figure below illustrates the training, validation and loss curves of the model.

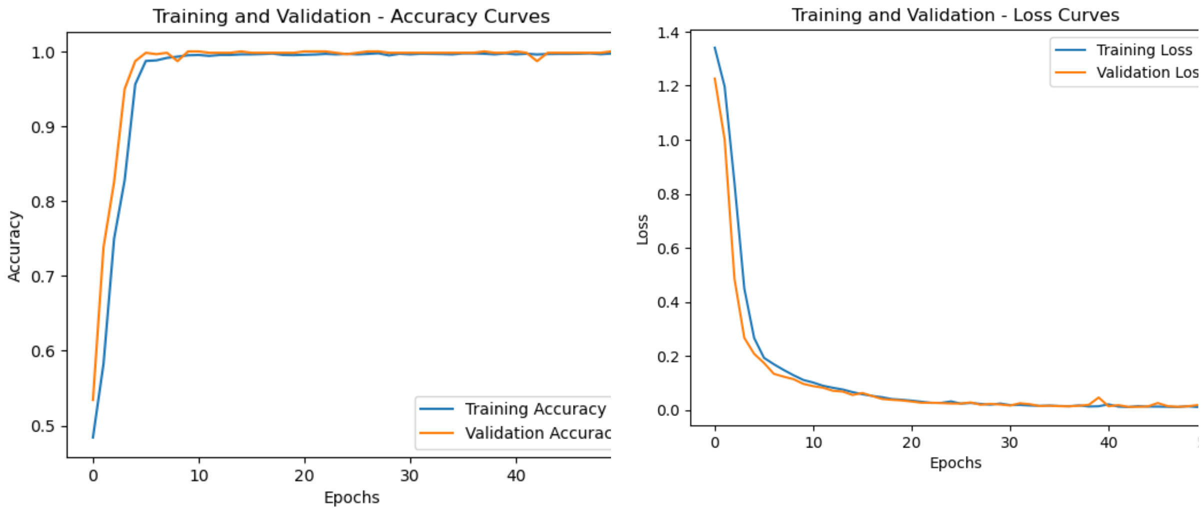

Figure 6-12 Accuracy and Loss Curve of FNN_LSTM Model Through Epochs

6.6 Summary of Model's Accuracy Scores

Table 6-1 Summary of Models Score

Stratified 10-fold Accuracy result (%) for Propose Models											
Models	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Accu (%)
BiLSTM	95.63	98.77	97.63	99.44	97.44	96.07	95.84	93.81	95.96	98.74	96.99
FNN	94.83	98.52	94.98	96.85	92.51	93.86	97.96	95.63	93.78	95.80	95.47
FNN_BiLSTM	98.74	99.81	98.70	99.45	98.88	98.76	97.85	99.42	98.83	98.68	98.94
FNN_LSTM	88.72	89.59	95.34	88.49	92.30	90.19	93.80	95.82	93.83	96.75	92.49
LSTM	98.72	98.5	98.14	98.99	98.58	96.38	97.61	97.07	93.03	97.94	97.49

In the above shown table presents the stratified 10-fold cross validation accuracy of each fold for BiLSTM, FNN, FNN_BiLSTM, FNN_LSTM and LSTM models. From these models the hybrid model of FNN with BiLSTM achieved a highest accuracy score of 98.94 compared with

FNN 95.47, LSTM 97.49, BiLSTM 96.99 and FNN_LSTM 92.49 respectively. Evaluation of Prediction metrics are presented in the below table.

6.7 Stratified 10-Fold Score of Classification Evaluation Metrics

Table 6-2 Stratified 10-fold Score of Classification Evaluation Metrics

Models	Precision (%)	Recall (%)	F1 score (%)
BiLSTM	0.97	0.97	0.96
FNN	0.96	0.95	0.95
FNN_BiLSTM	0.99	0.98	0.98
FNN_LSTM	0.95	0.92	0.90
LSTM	0.98	0.97	0.96

In the above table the result of different evaluation metrics is presented. From the above table FNN_BiLSTM outperformed the rest models with 0.99, 0.98 and 0.98 in precision, recall and F1-score respectively.

Figure 6-13 Comparison of Proposed Model

We have conducted and assessed a performance comparison between our model and two other machine learning models: a support vector machine and a random forest. The performance evaluation is detailed in Table 2, demonstrating that our model surpasses both of these algorithms in terms of average accuracy, precision, recall, and F1 score.

Table 6-3 Comparison of the proposed mode, RF and SVM

	Our Model (FNN_BiLSTM)	RF	SVM
Accuracy	98.94	97.40	89.48
Precision	98	97	81
Recall	98	97	89
F1-Score	98	97	85

6.8 Activation Function and Optimization Function

Deep learning is the subfield of machine learning which is used to perform complex tasks such as speech recognition, text Prediction. The complexity of the deep learning model is determined by its components, including activation functions, input and output layers, hidden layers, loss functions, and various other parameters. Every deep learning algorithm strives to create generalizations from the data it's exposed to, with the goal of making predictions on unseen data. Optimizer algorithms are methods for optimizing deep learning models, and they have a significant impact on both the accuracy and training speed of the model (Gupta, 2023).

An optimizer is a method that adjusts the characteristics of a neural network, such as its weights and learning rates. This leads to a reduction in overall loss and an improvement in accuracy. Because deep learning models typically consist of millions of parameters, determining the correct weights for the model can be a challenging task (Gupta, 2023). In this study the performance of 3 optimization function has been used in the proposed deep learning model. The performance of these 3 optimization functions such as Adam, RMSprop and Stochastic Gradient descent has been compared with each other to show the outperforming optimization function according to their performance in their respective deep learning models. The result of the above stated activation and optimization values are under 10-fold stratified cross validation. FNN_BiLSTM outperformed the rest deep learning approaches for this proposed study.

6.8.1 Hyperparameter Tunning

Hyperparameters are numerical or Boolean values that can be adjusted and configured by a user prior to the classifier training process. They play a crucial role in enhancing the model's training efficiency, performance, and predictive capabilities. To choose the optimum hyperparameter for the proposed RNN based models, this study has used grid search tunning techniques. Grid search is a method that finds the best hyperparameters for a given model through a thorough and exhaustive exploration of the available hyperparameter options. Each hyperparameter affects the model learning process (Hisham Allahem, 2022).

The selection of hyperparameter value ranges in the context of our study takes a specific significance. First and foremost, expertise from the medical domain becomes paramount, as understanding the distinctive risk factors, clinical features, and the importance of accurate Prediction is crucial. The investigation of prior research in the field, particularly studies related to healthcare data, can provide valuable insights into hyperparameter settings and model architectures that have shown promise in similar clinical applications. Secondly, Deep exploration of the tabular dataset is essential, considering patient-related features, medical history, prenatal care details, and other pertinent factors. Detecting potential class imbalances or data irregularitiesthat could affect hyperparameter choices is imperative. Additionally, balancing the intricacy of themodel with dataset size was taken into consideration during the selection of hyperparameters. These selection of hyperparameter ranges, such as the depth of the neural network, the number of neuronsper layer, learning rates, and batch sizes, were tailored to capture the intricate relationships withinmedical data while mitigating overfitting risks. In essence, the process of choosing hyperparametervalues becomes a highly specialized and meticulous endeavor, aligning the deep learning model'sconfiguration with the unique requirements and challenges of classifying adverse pregnancy outcomes in a medical context. From grid search experiments, the model has shown a better performance in terms of accuracy, precision, recall, and F1 score based on the following parametercombinations.

This study has used hyperparameters like learning rate, Epochs, Batch size, optimizers, activation functions, kernel regulation rate and dropout rates. By searching for the best number of neurons, number of layers, number of epochs, number of batch size, learning rate, activation function, optimization and loss functions we can tune our models. This study presented the hyperparameter ranges for all the proposed models. For the number of layers (1,2,3,4), Number of neurons (16,32,64, 128), for activation functions (ReLU, sigmoid, SoftMax), for optimizers (Adam, RMSprop, SGD), loss function (Categorical Cross entropy, MSE), Epoch (20, 32, 50 100), batch size (16, 20, 32, 50) and learning rate (0.01, 0.001, 0.0001, 0.05). From grid search experiments, the models have shown a better accuracy result based the following parameter of Number of Hidden layers **[3]**, Number of Neurons **[128]**, Activation Function **[ReLU]**, Optimization **[Adam]** for hidden layer and **[SoftMax]** for the output layer**,** Loss Function **[Categorical Cros entropy]**, Batch size **[32]**, Epoch **[50]** and Learning rate **[1e-3]** rate respectively.

6.9 Discussion

One of the aspects of this study is to build a deep learning-based model that predicts adverse pregnancy outcomes. To our knowledge, this is the first study to explore adverse pregnancy outcome Prediction based on data gathered from EDHS using deep-learning approaches. This study undergoes several experimentations as stated in chapters 5 and 6 by using the EDHS dataset on five proposed deep learning models to improve their performance. These proposed models, BiLSTM, LSTM, FNN, FNN_LSTM, and FNN_BiLSTM have been used and evaluated in this study. The grid search hyperparameter tunning algorithm is used in this work, to search for the optimal hyperparameters.

As explained in Table 6.1, 6.2, and Figure 6.13 the comparison has been made between all proposed models, and the FNN_BiLSTM model outperformed all the other proposed models with 98.94% accuracy under stratified 10-fold cross-validation. Based on the nature, of our dataset containing time series measurements and sequence of events FNN_BiLSTM model is suited the most. The proposed FNN_BiLSTM model can preprocess and engineer our datasets effectively. Also, the proposed model can use past and future information in the current state when working with sequential data by learning relevant patterns from the data. This is the reason why the proposed FNN_BiLSTM model outperformed the other proposed deep learning models. At the beginning of the study, three research questions were encompassed. This section will provide the answer to these questions raised in section 1.4 of the study.

RQ1. What are the most determining attributes that influence the occurrence of adverse pregnancy outcomes?

To answer this question, feature importance analysis was used to identify the most determinant attributes. We have also used gradient input and LASSO techniques to identify their feature contributions from their score. After we examined the performance of the model with the most contributed and least contributed features based on their feature contribution scores. From the findings in Figure 6.2, pregnancy complication is the most contributing feature which is followed by cigarette use as the second contributing feature followed by size of the child as the third contributing factor while anemia level is the least contributing factor. The result of the predictive model with the most contributing attribute has a great impact on the result while the performance of the model using the least contributing attribute was low in terms of accuracy, F1-score, and precision.

RQ2. Which deep learning model is best suited for developing a model for predicting adverse pregnancy outcomes?

For this question, the researcher has selected five models that will be suitable for the Prediction problem as a solution. These selected models are FNN, LSTM, BiLSTM, FNN_LSTM and FNN_BiLSTM. These selected models were trained, experimented, and tested using the prepared dataset. They were compared with each other by using various hyperparameter tunning techniques. To determine the best suited model for Prediction of adverse pregnancy outcomes, evaluation metrics, and Prediction metrics were used in the study. Based on the experiments FNN_BiLSTM outperforms the other proposed models with 0.9894% accuracy as the best classifier of adverse pregnancy outcome.

RQ3. What is the hyperparameter that has the most significant impact on the performance of the model during the experiment?

For this research question, various hyperparameter tunings were addressed by using grid search tunning techniques for all the proposed deep learning models. The first thing we did was we identified the hyperparameters that are best suited for our prediction problems. After doing that, we selected those parameters that work well with multi-class Predictions. The hybrid usage of FNN with BiLSTM with a set of parameters was experimented with to evaluate the performance of the model. Stratified cross-evaluation and Prediction metrics evaluation were used to evaluate the performance of the proposed model based on the set of parameters. The performance of the proposed models was tracked thoroughly during the experimentation and the optimum hyperparameter set was identified. We improved the performance of the model for adverse pregnancy outcomes by combining all of these factors.

Various studies that have been conducted previously on adverse pregnancy outcomes using machine learning and deep learning techniques are stated in Chapter Two Table 2.2. The studies mentioned below were compared to the newly proposed model as follows:

The first study was conducted by (Yu Mu, 2018) to predict adverse pregnancy outcomes using deep learning approaches with the help of MLNN. They have used MLNN to develop their model. In the study, they used a large number of feature sets that make feature engineering and selection challenging. Because managing and selecting relevant features from a large pool can be difficult. These affected the performance of the model negatively. In addition, optimization techniques were not incorporated in the study to the fullest.

The second study was conducted by (Hisham Allahem, 2022) to mitigate the consequence of premature birth for pregnant women and fetuses using ML and DL. The study used only one feature to address the problem. Important evaluation metrics like FPR and FNR were not included. As the study, is a binary Prediction problem they proposed a framework, not a predictive model. The number of datasets used for the purpose of the study was from 5 databases (604 instances) is small and they need to incorporate a large data size.

Other studies like those (Yandamur, 2018) were conducted to anticipate the pregnancy results of women who have Systemic Lupus Erythematosus during their pregnancy. This study focused on SLE-affected pregnant women only. They have not clearly stated the features they used for their study. As their problem is a binary Prediction, they have not incorporated contributing features in their study. Also (Ayleen Bertini, 2022) in their study tried to identify the applicable performance of ML approaches to predict pregnancy outcomes. Due to the fact that the study has used separate complications, baselines, and variable input, it is challenging to access and compare the prediction because some of the predictions lack variable measurements.

Moreover, the studies conducted previously on adverse pregnancy outcomes have gaps. From these gaps, the major ones are lack of depth in exploring the relationship among contributing attributes, lack of data pertaining to the determinant factors, lack of features, and results not being clearly stated in their studies are among the gaps that are mentioned in the above papers.

Some of the studies applied a binary Prediction problem which only works on two outcomes. In addition, the usage of a small number of datasets has also impacted the performance of the predictive model. The abovementioned problems significantly impacted the models' performance in terms of their respective Prediction metrics accounting for a lower number of accuracy scores.

The study we have proposed used various deep-learning approaches and preprocessing techniques to predict adverse pregnancy outcomes. The study target Ethiopian women by using data collected from EDHS. For this work, we have preprocessed and prepared our dataset following DHS program guidelines. These guidelines contain different maternal health categories including determinant features such as socio-demographic, obstetric, birth history, pregnancy history, nutritional, and other sections were included. Then we prepared our dataset for the model development. We have experimented using five different deep-learning algorithms for the prediction. These algorithms are BiLSTM, FNN, LSTM, FNN_LSTM and FNN_BiLSTM. The findings from the experiments are shown in Table 6.1,6.2, and Figure 6.1. We conclude the FNN_BiLSTM model outperformed the other proposed models by achieving an accuracy of 98.94% and 99% of F1-Score. To identify the best-suited parameters for the models, a grid search was carried out throughout the study. Besides Prediction metrics including accuracy, precision, recall, F1-Score, and confusion matrix, stratified 10-fold cross-validation was used to evaluate the performance of the model. This study is the first to our knowledge in our country using the EDHS dataset.

6.10 Contribution

This research work contributes in several ways to maternal health care in predicting adverse pregnancy outcomes. The following list are the summarized contributions of the study.

- Classifying adverse pregnancy outcomes: using the predictive model based on deep learning models, the study predicted three classes of adverse pregnancy outcomes from the prepared dataset.
- Improving the performance of the prediction: First and foremost, the most contributing attributes were identified and a serious of experimentation were conducted using deep learning models based on the prepared dataset for the model. Finally, the performance

of the model was measured and finetuned using evaluation metrics and hyperparameter tunning techniques respectively.

- Dataset Preparation and Exploration: exploring and preparing of a dataset while using deep learning approaches is the most important techniques. The new dataset was constructed from the data gathered from the EDHS.

CHAPTER SEVEN

7. CONCLUSION AND RECOMMENDATION

7.1 Conclusion

Adverse pregnancy outcomes are major public issues that are a leading cause of neonatal morbidity and mortality throughout the world in developing countries. Preterm birth is a global issue for the pregnant women and babies that can lead to death and life-long serious health problem. Low birth weight is a major concern in developing countries like Ethiopia. The study presents adverse pregnancy outcome Prediction using deep learning approaches. Different researches that are related with adverse pregnancy outcome Prediction using deep learning and machine learning approaches were studied thoroughly in the study. To address the objective of this study, the dataset was gathered from Ethiopian Demographic Health Survey collected from Ethiopian Statistics Agency in 2000, 2005, 2011, 2016 and mini-EDHS of 2018 within 5 years intervals. The records were reviewed by EPHI public health data researcher and reviewer, Midwifery expert who has specialization in the field of maternal health and pregnancy outcomes. To label adverse pregnancy outcomes as Low Birth Weight and Preterm Birth were decided by reviewing the dataset with the domain experts through communications. The study classifies three adverse pregnancy outcomes which are Low Birth Weight, Preterm Birth and Normal Birth. The dataset extracted from EDHS has a class imbalance and missing values. The collected data was preprocessed by using various preprocessing techniques like handling missing values, removal of unwanted features and categorical values and the likes were used to provide a well-organized and clean dataset for the proposed deep learning model. The preprocessed dataset has 10220 instances with 15 columns including the target label class. The study presented 5 deep learning models to predict adverse pregnancy outcome such as FNN, LSTM, BiLSTM, FNN_LSTM and FNN_BiLSTM. To conduct this study, we have done a total of 5 experiments. The study also presented the hyperparameter tunning using grid search tunning technique to improve the performance of the model. We have also used MLP for feature extraction to reduce the dimensionality of the dataset.

To improve the performance of the FNN, LSTM, BiLSTM, FNN_LSTM and FNN_BiLSTM models optimum hyperparameters were using by applying grid search tuning techniques. In this study from the proposed models, FNN_BiLSTM achieved an accuracy of 98.94 %, precision 99%, Recall 98% and 98% of F1-Score under stratifies 10-fold cross validation. We have also identified the most determinant attributes that contribute to the occurrence of adverse pregnancy outcome with the help of assistance from the experts of the medical filed.

7.2 Future Work

Finally, we recommend future researchers conduct a predictive model for pregnant women that predicts other adverse pregnancy outcomes like Abortion, Birth Defect and still birth. This study only includes preterm birth and low birth weight as adverse pregnancy outcome prediction. We also recommend a predictive model that can predict adverse pregnancy outcome presence among women from respective maternal determinant risk factors during pregnancy and over time. Furthermore, the research is open to any researchers to extend the study using EDHS dataset.

***

Special Acknowledgment

This research was sponsored by Adama Science and Technology University with a grant number

ASTU/SM-R/834/23

Adama, Ethiopia

***

REFERENCES

Abas AA, F. E. (2017). Knowledge about danger signs and symptoms of pregnant women attending antenatal care centers in Baghdad City. . *J Nurs Health Sci*, 6(4): 37–40.

Ahmed, Z. M. (2020). *Artificial Intelligence with multi-functional machine learning platform development for better healthcare and precision medicine.* Oxford University Press.

Al-Hindi, M. Y. (2020). Association of antenatal risk score with maternal and neonatal mortality and morbidity. . *Cureus*, 12, e12230. doi: 10.7759/cureus.12230.

Allen, J. L. (2021). Deep learning-based photoplethysmography Prediction for peripheral arterial disease detection: a proof-of-concept study.

Alshahrani, S. a. (2016). Are Deep Learning Approaches Suitable for Natural Language Processing? . *NLDB 2016: 21st International Conference on Applications of Natural Language to Information Systems.* .

Andreea M. Oprescu, G. M.-A.-D.-T. (2020, October 14). Artificial Intelligence in Pregnancy: A Scooping Review. *IEEE Acess*, 35.

Ayleen Bertini, R. S. (2022). Using Machine Learning to predict complication in pregnancy : A systematic review . *Frontiers in Bioengineering and Biotechnology*, doi: 10.3389/fbioe.2021.780389.

Belayneh Endalamaw D, T. M. (2022). Predicting the level of anemia among Ethiopian women using homogeneous ensemble machine learning algorithm. *BMC Medical Informatics and Decision Making* , https://doi.org/10.1186/s12911-022-01992-6.

Bhatt, A. R. (2021). Cervical cancer detection in pap smear whole slide images using convnet with transfer learning and progressive resizing. . *PeerJ Comput. Sci. 7,*.

Bhattacharya, S. C. (2007). Effect of Body Mass Index on pregnancy outcomes in nulliparous women delivering singleton babies. . *BMC Public Health 7,168*, https://doi.org/10.1186/1471-2458-7-168.

Centers for Disease Control and Prevention. (2008.). *Update on Overall Prevalence of Major Birth Defects–, 1978-2005.* Atlanta, Georgia: MMWR Morb Mortal Wkly Rep.

Chang Su, Z. X. (2020). Deep learning in mental health outcome research:. *Translational Psychiatry* .

Cheng Gao, S. O. (2019). Deep Learning Predicts Extreme Preterm Birth from Electronic Health Records. *Journal of Biomedical Informatics*, https://doi.org/10.1016/j.jbi.2019.103334.

Crear-Perry, J. C.-d.-A. (2021). Social and structural determinants of health inequities in maternal health. J. . *Women's Health. 30*, , 230–5. doi: 10.1089/jwh.2020.8882.

Creswell, J. W. ((2014). Research design: Qualitative, quantitative and mixed methodsapproaches . *SAGE Publications(4th ed.). Thousand Oaks:* .

CSA, “. s. (2016). *Ethiop. Demogr. Heal. Surv. Addis Ababa, Ethiop.* . Calverton, Maryland, USA, .

De Vaus, D. A. (2001). *Research design in social research.* London: SAGE,.

Dhiraj Dahiwade, P. G. (2019). Designing Disease Prediction Model Using Using Machine Learning. *International Conference on Computing Methodologies and Communication*.

Diguisto C, D. V. (2020). Consequences of active cigarette smoking in pregnancy - CNGOF-SFT expert report and Guidelines on the management of smoking during pregnancy. . *Gynecol Obstet Fertil Senol.* , 48(7–8):559–66.

Dithy M. D, D. K. (2022). Predicting anemia in pregnant women by using gausnominal Prediction algorithm. *International Journal of Pure and Applied Mathematics*, ISSN: 1314-3395 (on-line version) .

Doherty DA, M. E. (2006). re-pregnancy body mass index and pregnancy outcomes. *Int J Gynaecol Obstet.* , ;95:242–247.

Finlayson K, C. N. (2020). What matters to women in the postnatal period: a meta-synthesis of qualitative studies. . *PloS one.*, 15(4):e0231415. pmid:32320424.

Fowler, J. R., Mahdy, H., Jack, & W., B. (2023). Pregnancy. *National Library of Medicine*.

Gebresilassie, B. B. (2019). Timing of first antenatal care attendance and associated factors among pregnant women in public health institutions of Axum town, Tigray, Ethiopia, 2017: a mixed design study. *BMC Pregnancy Childbirth*, https://doi.org/10.1186/s12884-019-2490-5.

Girma, S. F. (2019). Factors associated with low birthweight among newborns delivered at public health facilities of Nekemte town, West Ethiopia: a case control study. *BMC Pregnancy Childbirth 19*, . https://doi.org/10.1186/s12.

Gregory ECW, V. C. (2022.). *Fetal mortality:* . United States, Hyattsville, MD: National Center for Health Statistics. 2022. National Vital Statistics Reports; vol 71 no 4. .

Gupta, A. (2023). A Comprehensive Guide on Optimizers in Deep Learning. *analyticsvidhya*.

Habyarimana F, Z. T. (2020). Prevalence and risk factors associated with anemia among women of childbearing age in Rwanda. *Afr JReprod Health.*, 24(2):141–51. https://doi.org/10.29063/ajrh2020/.

Hailu M, G. A. (2010). Knowledge about obstetric danger signs among pregnant women in AlettaWondo district, Sidama zone, Southern Ethiopia. *Ethiop J Health Sci.*, 20(1):25–32.

Hasan Rawashdeh, S. A. (2019). Intelligent system based on data mining techniques for prediction of preterm birth for women with cervical cerclage. *Computational Biology and Chamistry* , https://doi.org/10.1016/j.compbiolchem.2020.107233.

Hisham Allahem, S. S. (2022). Automated labour detection framework to monitor pregnant women with a high risk of premature labour using machine learning and deep learning . *Informatics in Medicine Unlocked*, https://doi.org/10.1016/j.imu.2021.100771.

Hornstra G, U. R. (2005). The impact of maternal nutrition on the offspring. . *Basel: Karger Medical and Scientific Publishers;* .

Hug L, M. A. (2020). *A Neglected Tragedy. The Global Burden of Stillbirths:* . New York, NY: United Nations Children's Fund (2020).: Report of the UN Inter-agency Group for Child Mortality Estimation,.

Islam, M. S. (2018). A Deep Recurrent Neural Network with BiLSTM model for Sentiment Prediction. .

Jogin, M. &. (2018). Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning. . *Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2319-2323. .

Jour Schwenker, F. C. (2019). A Stacked BiLSTM Neural Network Based on Coattention Mechanism for Question Answering ,. *SP*, 9543490.

Jovancic, N. (2020.). Research design types key elements andcharacteristics.

Junyi Chai, H. Z. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications,*, Volume 6,.

Kamineni V, M. A. (2017). Birth preparedness and complication readiness in pregnant women attending urban tertiary care hospital. . *J Family Med Prim Care.* , 297-300 doi: 10.4103/2249-4863.220006. PMID: 29302536; PMCID: PMC5749075.

Kataoka MC, C. A. (2018). Smoking during pregnancy and harm reduction in birth weight: a cross-sectional study. . *BMC Pregnancy Childbirth.*, 18(1):67.

Kavsaotlu AR, P. K. (2015). Non-invasive prediction of hemoglobin level using machine learning techniques with the PPG signal's. *Appl Soft Comput J.*, ;37:983–91.

Kaye DK, M. F. (2006). Domestic violence during pregnancy and risk of low birth weight and maternal complications: a prospective cohort study at Mulago Hospital, Uganda. . *Trop Med Int Health.*, 11:1576–84. .

L. Macías-García, J. L.-R.-G.-B.-S.-C. (2017). A study of the suitability of autoencoders for preprocessing data in breast cancer experimentation. *J. Biomed. Inform., 72* , 33-44.

Lale, S. C.-B. (2015). *Global causes of maternal death: a WHO systematic analysis.* The Lancet Global Health.

Lawn JE, B. H. (2016). Stillbirths: rates, risk factors, and acceleration towards 2030. *Lancet.* , 387(10018):587–603.

Lawn JE, L. A. (2009). Two million intrapartum-related stillbirths and neonatal deaths: where, why, and what can be done? *Int J Gynecol Obstet.* , 107:S5–19.

Lee AC, e. a. (2013). National and regional estimates of term and preterm babies born small for gestational age in 138 low-income and middle-income countries in 2010. *Lancet Glob Health.* , 26–36.

Li WU, Y.-l. S.-y. (2020). Associations of maternal height with adverse birth outcomes. . *Chinese Journal of Public Health,* , 1318-1322. doi: 10.11847/zgggws1123794.

Londero, A. P. (2019). Maternal age and the risk of adverse pregnancy outcomes: a retrospective cohort study. . *BMC Preg. Childbirth*, doi: 10.1186/s12884-019-2400-x.

Magnus MC, W. A. (2019). Role of maternal age and pregnancy history in risk of miscarriage: Prospective register based study. :BMJ 364:l869, doi: 10.1136/bmj.l869.

Mardiyanti I, N. D. (2019). The independence of pregnant women in early detection of high risk of pregnancy terms of parity and information exposure. *J Public Health Africa*.

McClure EM, S. S.-T. (2015). Stillbirth rates in low-middle income countries 2010 - 2013: a population-based, multi-country study from the Global Network. *Reprod Health.*, doi: 10.1186/1742-4755-12-S2-S7.

Mehari, M. M. (2020.). Advanced maternal age pregnancy and its adverse obstetrical and perinatal outcomes in Ayder comprehensive specialized hospital, Northern Ethiopia, 2017: a comparative cross-sectional study. . *BMC Pregnancy Childbirth 20,* , https://doi.org/10.1186/s12884-020-2740-6.

Melissa J Fazzari, M. M. (2022). Adverse pregnancy outcomes in women in women with systemic lupus erythematosus: can we improve predictions using machine learning? *Lupus Science & Medicine*, doi:10.1136/lupus-2022-000769.

Mesfin Tadese, K. D. (2022). Assessment of the adverse pregnancy outcomes and its associated factors among deliveries at Debre Berhan Comprehensive Specialized Hospital, Northeast Ethiopia. *PLoS ONE*, https://doi.org/10.1371%2Fjournal.pone.0271287.

Miotto, R. W. (2018). Deep learning for healthcare: review, opportunities and challenges. . *Briefings Bioinform. 19, 1236–1246*, doi: 10.1093/bib/bbx044.

Monroy De Velasco, A. (1982). Consequences of early childbearing. *Draper Fund Rep.*, 26-7.

Muchie KF, e. a. (2020). Epidemiology of preterm birth in Ethiopia: systematic review and meta-analysis. *BMC Pregnancy Childbirth*, 20(1):1–12.

Muhammad Nazrul Islam, S. N. (2022). Machine Learning to predict pregnancy outcomes : a systematic review , sythesizing framework and future research agenda. *BMC Pregnancy and Childbirth* , https://doi.org/10.1186/s12884-022-04594-2.

Mullick, S. (2005). Sexually transmitted infections in pregnancy: prevalence, impact on pregnancy outcomes, and approach to treatment in developing countries. *Sexually Transmitted Infections* , https://doi.org/10.1136/sti.2002.004077.

Orginization, W. H. (2018). WHO Recommendations on Antenatal Care for a Positive Pregnancy Experinece: Summary . *Maternal and Survival Helath program* .

Ramadhan, M. &. (2017). Parameter Tuning in Random Forest Based on Grid Search Method for Gender Prediction Based on Voice Frequency. . *DEStech Transactions on Computer Science and Engineer*.

Ritchie., M. R. (2017). Maternal Mortality - Our world in Data. https://ourworldindata.org/maternal-mortality.

Rout AR, B. S. (2018). Natural scene Prediction using deep learning. In: 2017 international conference on computing, communication, control and automation,. *ICCUBEA 2017*, http://dx.doi.org/10.1109/ICCUBEA.2017.8463727.

S. Chen, H. X. (2014). A vision of IoT: applications, challenges, and opportunities with china perspective. *IEEE Internet Things J.*,, 349-359.

S. Min, B. L. (2017). Deep learning in bioinformatics. *Briefings Bioinformatics*, 851-869.

Sashikanta Prusty, S. K. (2022). SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer. *Front. Nanotechnol.*, .

Services, O. o. (2021, Feburary 22). *Office on women's health.* Retrieved from Office on women's health: https://www.womenshealth.gov/pregnancy/youre-pregnant-now-what/stages-pregnancy

Shahab Shamshirband, M. F.-R. (2021). A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues,. *Journal of Biomedical Informatics,*, Volume 113, https://doi.org/10.1016/j.jbi.2020.103627.

Stephen G, M. M.-P. (2018). Prevelance , eisk factors and adverse perinatal outcomes in northern tanzania. *Anemia inpregnancy*.

System, J. H. (2021). *Jackson on health.* Retrieved from Jackson on health : https://maternity.jacksonhealth.org/pregnancy-milestones/

Tadese M, D. K. (2022). Assesment of the advesre pregnancy outcomes and its factirs among deliveries at Debre Berhan Comprehensive Specialized Hospital, Northest Ethiopia. *PLoS ONE*, 17(7): e0271287. https://doi.org/10.1371/journal.pone.0271287.

Tafere Birlie Ayele, Y. A. (2023). Prevalence of preterm birth and associated factors among mothers who gave birth in public hospitals of east Gojjam zone, Ethiopia. *BMC Pregnancy and Childbirth*.

Tobore I, L. J.-H. (2019). Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations. *JMIR Mhealth Uhealth* , https://doi.org/10.2196/11966.

Unger, H. T. (2019). The assessment of gestational age: a comparison of different methods from a malaria pregnancy cohort in sub-Saharan Africa. . *BMC Pregnancy Childbirth 19*, , https://doi.org/10.1186/s12884-018-2128-z.

van der Graaf, R. v. (2018). Fair inclusion of pregnant women in clinical trials: an integrated scientific and ethical approach. . Trials. 19, 78. doi.

W. Sun, B. Z. (2017). Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis. *Comput. Biol. Med*, 530-539.

Wassihun B, N. B. (2020). Knowledge of obstetric danger signs and associated factors: a study among mothers in shashemene town, oromia region, Ethiopia. *Repord Health* .

WHO. (2019). Maternal mortality ratio (modeled estimate, per 100,000 live births) |Data;. https://data.worldbank.org/indicator/SH.STA.MMRT?end=2017&start=2000&view=chart.

WHO. (2021, Jan 10). Retrieved from Maternal deaths decline slowly with vast inequalities worldwide.: https://www.who.int/news/item/19-09-2019-maternaldeaths-decline-slowly-with-vast-inequalities-worldwide

WHO, U. ,. (2019). *Trends in maternal mortality : 2000 to 2017* . Geneva: World Health Organization .

WHO, U. U. (2016). *Trends in maternal mortality: 1990 to 2010.* . WHO, UNICEF, UNFPA, and The World Bank Estimates, .

WHO, U. U. (2019). *Trends in maternal mortality*. Geneva: World Health Organization. Retrieved from https://fragilestatesindex.org/

WHO. (2006). *Standards for maternal and neonatal care: birth and emergency preparedness in antenatal care: department of making pregnancy safer (MPS).* Geneva:: World Health Organization.

Xang S, Z. F. (2021). Intelligent Health Care : Application of Deep Learning in Computational Medicine. *frontier in genetics*, 21.

Yandamur, H. (2018). Predicting The Pregnancy Outcome in SLE Affected Women Using Recurrent Neural Network . *IJRAR* .

Yeshialem E., A. M. (2019). “Determinants of adverse pregnancy outcomes among mothers who gave birth from jan 1-dec 31/2015 in jimma university specialized hospital, case control study, 2016,” . *Ethiop. J. Reprod. Heal.*, https://doi.org/10.1371/journal.pone.0271287.

Yu Mu, K. F. (2018). Applying deep learning for adverse pregnancy outcome detection with pre-pregnancy health data. *MATEC Web of Conferences*, https://doi.org/10.1051/matecconf/201818910014.

APPENDICES

Appendix A: Sample Dataset

V228	M18$1	M45$1	V437	V438	V457	M14$1	V445	V763A	D106	V447A	V463A	V233	M15$1	outcome
1	5	0	38.4	1.531	4	2	16.38	0	0	24	0	4	11	1
0	3	0	52.2	1.591	4	1	20.62	0	0	31	0	4	11	0
1	5	0	40	1.51	4	1	17.54	0	0	43	0	4	11	1
0	3	0	51	1.554	4	3	21.12	0	0	36	0	4	11	0
0	3	0	57.4	1.697	4	7	19.93	0	0	30	0	4	22	0
0	3	1	63.6	1.607	4	6	24.63	0	0	30	0	4	22	0
0	3	0	65	1.642	4	6	24.11	0	0	35	0	4	22	0
0	5	0	49	1.582	3	9	19.58	0	0	25	0	4	22	0
1	5	0	58.6	1.495	4	3	16.22	1	1	17	1	7	12	2
0	1	0	55	1.578	2	3	22.09	0	0	23	0	4	11	0
0	5	0	36.9	1.496	4	3	16.49	0	0	25	0	4	11	0
0	1	0	44	1.536	4	4	18.65	0	0	20	0	4	21	0
0	1	1	69.3	1.703	4	4	23.89	0	0	30	0	4	11	0
0	4	0	57.8	1.553	4	5	23.97	0	0	36	0	4	21	0
1	2	0	55.2	1.572	4	1	17.34	1	1	18	1	8	96	2
0	1	0	64.1	1.608	4	6	24.79	0	0	33	0	4	21	0
0	4	0	42.4	1.574	4	7	17.11	0	0	25	0	4	21	0
1	4	0	47.3	1.499	3	1	16.05	0	0	30	0	4	11	1
1	4	0	48.6	1.573	4	2	17.64	0	0	26	0	4	11	1
1	5	0	44.5	1.519	4	2	18.44	0	0	35	0	4	11	1
0	3	1	54	1.549	4	3	22.51	0	0	22	0	4	11	0
1	5	0	39.1	1.434	3	1	14.64	1	0	16	0	4	11	1
0	3	0	48.5	1.593	3	4	19.11	0	0	20	0	4	11	0
1	3	0	49.4	1.513	4	1	28.58	1	1	44	1	7	11	2
1	5	0	50.1	1.596	4	2	17.67	1	1	18	1	9	11	2
0	3	1	66.2	1.668	4	5	23.79	0	0	26	0	4	24	0

Appendix B: Sample Code for Proposed Models

FNN_BiLSTM Model sample code

```
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Flatten, Dropout, Bidirectional, LSTM
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
import pandas as pd
from sklearn.model_selection import StratifiedKFold
import matplotlib.pyplot as plt
import seaborn as sns
from tensorflow.keras import regularizers
# Load and preprocess the dataset
dataset = pd.read_csv('C:/Users/Fozi/Desktop/BCD/RefinedBook1.csv')
X = dataset.iloc[:, 0:14]
y = dataset.iloc[:, 14]
n_splits = 10
kfold = StratifiedKFold(n_splits=n_splits, shuffle=True, random_state=42)
for i, (train_indices, val_indices) in enumerate(kfold.split(X, y)):
    print(f"Fold {i+1}/{n_splits}")
# Split the data into train and validation sets for the current fold
X_train, X_val = X.iloc[train_indices], X.iloc[val_indices]
y_train, y_val = y_categorical[train_indices], y_categorical[val_indices]
# Convert the data to numpy arrays
X_train = X_train.values.reshape((-1, 14, 1))
X_val = X_val.values.reshape((-1, 14, 1))
# FNN with LSTM model
fnn_bilstm_model = Sequential()
fnn_bilstm_model.add(Bidirectional(LSTM(128, input_shape=(14, 1), activation='relu', kernel_regularizer=regularizers.l2(0.
fnn_bilstm_model.add(Dropout(0.8))  # Add dropout layer
fnn_bilstm_model.add(Dense(128, activation='relu', kernel_regularizer=regularizers.l2(0.08)))
fnn_bilstm_model.add(Dropout(0.8))  # Add dropout layer
fnn_bilstm_model.add(Dense(3, activation='sigmoid', kernel_regularizer=regularizers.l2(0.08)))
fnn_bilstm_model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
# Train the model
history = fnn_bilstm_model.fit(X_train, y_train, validation_data=(X_val, y_val),epochs=50, batch_size=32)
# Evaluate the model on the validation set
y_val_pred =  fnn_bilstm_model.predict(X_val)
y_val_pred_classes = tf.argmax(y_val_pred, axis=1)
y_val_true_classes = tf.argmax(y_val, axis=1)
accuracy = accuracy_score(y_val_true_classes, y_val_pred_classes)
precision = precision_score(y_val_true_classes, y_val_pred_classes, average='weighted')
recall = recall_score(y_val_true_classes, y_val_pred_classes, average='weighted')
```

Result of SVM and RF model

```
Support Vector Machine (SVM) Model
Average Accuracy: 89.48%
Average Precision: 0.81
Average Recall: 0.89
Average F1 Score: 0.85

Random Forest Model
Average Accuracy: 97.40%
Average Precision: 0.97
Average Recall: 0.97
Average F1 Score: 0.97
```