**PREDECTING** **MATETNAL HEALTH RISK USING COMPOSITE HYPERCUBES ON ITERATED RANDOM PROJECTIONS ALGORITHM**

MD. Bappy Islam

Student Id:20191005010

Morshed Arifin

Student Id:20191010010

Sripurna Bachhar

Student Id:20191033010

Sneha mim

Student Id:20191044010



Department Of Computer Science and Engineering

Faculty of Science and Technology

North Western University, Khulna-9000, Bangladesh.

September-202

**Certificate**

This is to certify that the thesis entitled **“PREDECTING MATETNAL HEALTH RISK USING COMPOSITE HYPERCUBES ON ITERATED RANDOM PROJECTIONS ALGORITHM”** has been submitted by 20191005010, 20191010010,20191033010,20191044010 is partial fulfillment for the requirement of the degree of B.Sc. Engineering in Computer Science and Engineering, under the faculty of the Science and Technology, North Western University, Khulna-9000, Bangladesh.

Authors

…………………………………

MD. Bappy Islam

Id No:20191005010

………………………………….

Morshed Arifin

Id No:20191010010

………………………………….

Sripurna Bachhar

Id No:20191033010

……………………………………

Sneha mim

Id No:20191044010

**PREDICTING MATERNAL HEALTH RISK USING COMPOSITE HYPERCUBES ON ITERATED RANDOM PROJECTIONS ALGORITHM**

A thesis submitted in partial fulfillment of the requirement of the degree of Bachelor of

Science in Computer Science and Engineering.

-------------------------------

Nagib Mahfuz Supervisor Senior Lecturer

Department of Computer Science and Engineering, North Western University, Khulna Bangladesh.

-------------------------------

Md. Mahedi Hasan Second Supervisor

Senior Lecturer

Department of Computer Science and Engineering, North Western University, Khulna.

Bangladesh.

-------------------------------

Tajul Islam Head of the Department

Assistant Professor

Department of Computer Science and Engineering, North Western University, Khulna.

Bangladesh.

**Abstract**

Maternal health is a critical aspect of public health, and accurately predicting health risks during pregnancy plays a vital role in preventing adverse outcomes for both the mother and the child. This paper introduces a novel approach to maternal health risk prediction by utilizing composite hypercubes on an Iterated Random Projections (CHIRP) algorithm.

The proposed method aims to address the challenge of effectively analyzing large-scale maternal health datasets characterized by high dimensionality and complex relationships between various risk factors. The CHIRP algorithm is employed as a dimensionality reduction technique, which facilitates the extraction of relevant features and reduces computational complexity. The key innovation lies in the integration of composite hypercubes, a powerful mathematical framework for modeling complex data structures, into the CHIRP algorithm. By mapping the multidimensional health data onto composite hypercubes, the algorithm captures the intricate interactions between different risk factors, enabling more accurate risk assessment. To evaluate the effectiveness of the proposed approach, a comprehensive analysis was conducted using a diverse dataset of maternal health records. The experimental results demonstrate that the composite hypercubes on CHIRP algorithm outperforms existing methods in predicting maternal health risks, achieving higher accuracy and improved interpretability. The findings of this study have significant implications for maternal health care providers and policymakers. By accurately predicting health risks during pregnancy, healthcare professionals can implement timely interventions and preventive measures to mitigate potential complications. Furthermore, the interpretability of the composite hypercubes on CHIRP algorithm enables healthcare practitioners to gain insights into the underlying factors contributing to specific health risks, facilitating personalized care and targeted interventions. In conclusion, this research introduces a novel approach for predicting maternal health risk using “composite hypercubes on the Iterated Random Projections” algorithm. The integration of composite hypercubes enhances the interpretability and predictive accuracy of the algorithm, enabling more effective decision-making in maternal health care. Future research can focus on expanding the application of this approach to other healthcare domains and refining the algorithm's performance through further optimization techniques.

**Keywords:** Maternal health risks, Data mining, CHIRP algorithm, Risk factors, pregnancy , health data, risk assessment, health care, Predictive accuracy, Personalized care

**Table of Contents**

[Chapter 1 6](#_Toc136028355)

[Introduction 6](#_Toc136028356)

[1.1 Background 6](#_Toc136028357)

[1.2 Motivation 6](#_Toc136028358)

[1.3 Contribution 7](#_Toc136028359)

[1.4 Objectives: 7](#_Toc136028360)

[1.4 Organization of the thesis: 8](#_Toc136028361)

[Chapter 2 9](#_Toc136028362)

[Background and Related Works 9](#_Toc136028363)

[2.1 Introduction 9](#_Toc136028364)

[2.2 Maternal health 9](#_Toc136028365)

[2.3 Maternal health risk 9](#_Toc136028366)

[2.4 Gestational Diabetes 10](#_Toc136028367)

[2.5 Preeclampsia 12](#_Toc136028368)

[2.6 Preterm Deliveries 13](#_Toc136028369)

[2.7 Recurrence and long-term maternal health risks of hypertensive disorders of pregnancy: a population-based study 15](#_Toc136028370)

[2.8 The maternal health clinic: an initiative for cardiovascular risk identification in women with pregnancy-related complications 16](#_Toc136028371)

[2.9: Maternal Healthcare in Migrant: A Systematic Review 17](#_Toc136028372)

[Chapter 3 19](#_Toc136028373)

[Literature Review 19](#_Toc136028374)

[3.1 Introduction 19](#_Toc136028375)

[3.2 Feature Details 19](#_Toc136028376)

# Chapter 1

## Introduction

### 1.1 Background

Maternal health refers to the health of women during pregnancy, childbirth and the postnatal period. Each stage should be a positive experience, ensuring women and their babies reach their full potential for health and well-being [1]. Maternal health is a critical aspect of public health, focusing on the well-being of pregnant women before, during, and after childbirth. Ensuring optimal maternal health is crucial not only for the mother's well-being but also for the health and development of the child. Unfortunately, pregnancy can pose various health risks to women, and accurately predicting and managing these risks is essential for preventing adverse outcomes. Accurately predicting health risks during pregnancy plays a vital role in proactive healthcare management. Early identification of potential complications allows healthcare professionals to implement appropriate interventions, monitor the progression of the pregnancy, and provide necessary support to mitigate risks. Additionally, predicting maternal health risks enables healthcare providers to offer personalized care tailored to the specific needs of each pregnant woman. In recent years, advancements in data analytics and machine learning techniques have provided new opportunities for predicting and managing maternal health risks. These approaches leverage large-scale maternal health datasets to identify patterns, correlations, and risk factors associated with adverse outcomes. By extracting valuable insights from these datasets, healthcare professionals can make more informed decisions and implement preventive measures [2]. One promising approach in the field of maternal health risk prediction is the utilization of composite hypercubes on the Iterated Random Projections (CHIRP) algorithm. Composite hypercubes provide a mathematical framework for modeling complex data structures, allowing for the capture of intricate interactions between different risk factors. The CHIRP algorithm, combined with composite hypercubes, enables dimensionality reduction and extraction of relevant features, thereby improving the accuracy and interpretability of risk predictions. [3]

### 1.2 Motivation

Accurately predicting maternal health risks during pregnancy is crucial for improving outcomes for both mothers and infants. This motivation stems from the desire to enhance maternal and neonatal well-being by identifying potential complications early on and implementing targeted interventions. Effective prediction models can prevent adverse outcomes such as preterm birth, gestational diabetes, and maternal infections, leading to better overall health for mothers and children. Additionally, accurate risk prediction allows healthcare providers to allocate resources efficiently, ensuring appropriate care for high-risk pregnancies. By leveraging large-scale datasets and advanced analytics, predictive models provide evidence-based insights for clinical decision-making and policy development in maternal health. Furthermore, integrating predictive analytics fosters strong patient-provider relationships by empowering expectant mothers with information about potential risks and enabling informed decision-making. Overall, the motivation for predicting maternal health risks lies in the pursuit of proactive and personalized care that can significantly impact the well-being of pregnant women and their infants. Here we have used Naive Bayes, J48, Multilayer Perceptron, Composite Hypercubes on Iterated Random Projections by supplying dataset instances according to our experimental environment.

### 1.3 Contribution

* Developed a novel approach for predicting maternal health risk using Composite Hypercubes on Iterated Random Projections (CHRIP) algorithm.
* Implemented CHRIP algorithm, which combines the power of Composite Hypercubes and Iterated Random Projections, to effectively predict maternal health risk.
* Conducted extensive research and analysis on a large dataset of maternal health records, considering various risk factors such as medical history, demographics, and lifestyle factors.
* Utilized Composite Hypercubes, a multidimensional data structure, to efficiently store and process the high-dimensional maternal health data.
* Employed Iterated Random Projections, a dimensionality reduction technique, to reduce the computational complexity and improve the prediction accuracy.
* Developed a comprehensive pipeline for preprocessing the maternal health data, including data cleaning, feature selection, and normalization, to ensure high-quality input for the CHRIP algorithm.
* Conducted rigorous experiments and evaluations on real-world maternal health data, comparing the performance of CHRIP algorithm with other state-of-the-art prediction models.
* Demonstrated that CHRIP algorithm outperforms existing approaches in terms of prediction accuracy, computational efficiency, and scalability, making it a promising tool for predicting maternal health risk.
* Provided insights and interpretations on the identified risk factors and their impact on maternal health outcomes, contributing to a better understanding of the underlying patterns and dynamics.
* Proposed potential applications of CHRIP algorithm in healthcare systems, such as early identification of high-risk pregnancies, personalized risk assessments, and targeted interventions for improving maternal health outcomes.
* Contributed to the field of predictive analytics in healthcare by introducing a novel algorithm that can enhance the effectiveness of maternal health risk prediction, potentially leading to improved healthcare decision-making and resource allocation.

### 1.4 Objectives:

The objective of "Predicting Maternal Health Risk Using Composite Hypercubes on Iterated Random Projections Algorithm" is to develop a method or system that can accurately predict maternal health risk using the Composite Hypercubes on Iterated Random Projections (CHIRP) algorithm. The specific objectives of this research or project may include:

* **Improve maternal health outcomes:** By accurately predicting maternal health risk, the aim is to improve the overall health outcomes for pregnant women and reduce the incidence of complications during pregnancy.
* **Early detection of high-risk pregnancies:** The objective is to identify pregnancies that are at a higher risk of complications or adverse outcomes at an early stage, allowing healthcare providers to intervene and provide appropriate care and support.
* **Personalized healthcare interventions:** The goal is to enable personalized healthcare interventions based on the predicted risk levels. By identifying high-risk pregnancies, healthcare providers can tailor their interventions and treatments to each individual, ensuring the best possible care.
* **Development of an efficient prediction algorithm:** The objective is to develop and optimize the CHIRP algorithm to accurately predict maternal health risk. This may involve refining the algorithm's parameters, enhancing its performance, and validating its accuracy through rigorous testing and evaluation.
* **Utilization of composite hypercubes:** The objective is to leverage the power of composite hypercubes in the prediction process. Composite hypercubes can provide a multidimensional representation of the data, allowing for a comprehensive analysis and identification of patterns or indicators of maternal health risk.
* **Contribution to research and knowledge:** The objective is to contribute to the existing body of research on maternal health risk prediction and enhance the understanding of the factors that influence maternal health outcomes. This research may lead to further advancements in the field and inform future studies.

### 1.4 Organization of the thesis:

The organization of the thesis is organized as follows:

**Chapter 2 (Related Works):**

Initiates about the background of predicting maternal health risks and the present several existing techniques machine learning, how to implement algorithms in their thesis, what are their advantages and limitations.

**Chapter 3 (Literature Review):**

Depicts the literatures and various related features and their working procedures that have been used in this study of predicting maternal health risks

**Chapter 4 (Methodology):**

Proposed our idea and detailed on the system design phase. The design phase includes the proposed techniques, features selection, classification and performance measurement.

**Chapter 5 (Simulation and Result Analysis):**

To validate the effectiveness of the proposed method and the outcome. In this chapter we will enclose the evaluation and results of our proposed techniques.

**Chapter 6 (Conclusion):**

Finally, we conclude our study and proposes some future works regarding these areas.

# Chapter 2

## Background and Related Works

### 2.1 Introduction

In this chapter the background of predicting maternal health risks and some related works of predicting maternal health risks are briefly described.

### 2.2 Maternal health

Maternal health refers to the health of women during pregnancy, childbirth and the postnatal period. Each stage should be a positive experience, ensuring women and their babies reach their full potential for health and well-being [1]. Maternal health is a critical aspect of public health, focusing on the well-being of pregnant women before, during, and after childbirth. Ensuring optimal maternal health is crucial not only for the mother's well-being but also for the health and development of the child

### 2.3 Maternal health risk

Maternal health risks refer to the potential dangers and complications that women may face during pregnancy, childbirth, and the postpartum period. These risks can have severe consequences for both the mother and the child. In many parts of the world, maternal health risks remain a significant public health concern, despite advances in medical knowledge and technology. [4]

Several factors contribute to maternal health risks, including inadequate access to healthcare services, poor nutrition, limited education, and socio-economic disparities. These challenges are particularly prominent in low-income and developing countries, where women may lack essential prenatal care, skilled birth attendants, and emergency obstetric services. As a result, complications such as hemorrhage, infections, hypertensive disorders, and obstructed labor can arise, leading to maternal morbidity and mortality. [5]

According to the World Health Organization (WHO), approximately 810 women die every day globally from preventable causes related to pregnancy and childbirth. Maternal health risks also extend beyond physical health, impacting women's overall well-being and quality of life. [4]

Addressing maternal health risks requires a comprehensive approach involving improved access to quality healthcare services, increased investment in maternal and reproductive health, and the empowerment of women through education and economic opportunities. Additionally, interventions such as early detection and management of complications, skilled birth attendance, and emergency obstetric care can significantly reduce maternal mortality rates.

Efforts to improve maternal health have shown positive results in many countries. However, sustained commitment and ongoing efforts are crucial to ensure that all women have the opportunity to experience safe and healthy pregnancies and childbirths.

### 2.4 Gestational Diabetes

Gestational diabetes mellitus (GDM) is defined as glucose intolerance with onset or first recognition during pregnancy. The definition does not require any return to normal glucose levels following delivery. Thus, GDM simply represents relatively high glucose levels at one point in the life of a young woman [6].

Outside of pregnancy, screening for clinically important levels of hyper glycaemia is generally recommended only for individuals with specific risk profiles. By contrast, screening for abnormal glucose levels is generally recommended as a routine component of care for pregnant women. Traditionally, screening during pregnancy has involved two steps. The first is a simple 1 h glucose challenge test to identify a large number of women at very low risk of clinically important hyper glycaemia; they do not need additional testing. The second step is a more complex 2 h or 3 h oral glucose tolerance test applied to the ‘at risk’ women to define the subset who have GDM. Specific cut-off points used in this detection process have varied widely. Relatively low cut-off points result in relatively high incidence rates of GDM, including many women with relatively mild hyper glycaemia. Relatively high cut-off points result in the converse.

For the purposes of this discussion, the specific cut-off points are less important than the general concept that GDM is diagnosed following a form of population screening for hyper glycaemia in young women. That screening occurs at a time when the women are generally quite insulin resistant, the acquired insulin resistance of late pregnancy might not be a dominant feature of the pathogenesis of GDM. the hyper glycaemia of GDM appears to have a small but demonstrable effect on perinatal outcomes, and is also associated with important long-term health problems in affected mothers and their children [6].

When women with history of gestational diabetes (GDM) undergo the 75-gram GTT at 6–12 weeks postpartum, 2–16% are diagnosed with type 2 diabetes (DM) and 36% are found to have intolerance to carbohydrates. Women who had prior GDM have a 36–70% risk of developing type 2 DM later in life, depending on risk factors and length of follow-up. It is important for women who had GDM to have appropriate follow up since, over time, often before patients are diagnosed, DM causes damage to various organs (heart, blood vessels, kidneys, eyes, nerves, etc.) Despite the deceptively benign term, “intolerance to carbohydrates”, this condition is associated with significant morbidity as well. Lee et al. performed a meta-analysis that included 15 prospective studies with 760,925 participants and reported that pre-diabetes, defined as impaired glucose tolerance or a combination of impaired fasting glucose and impaired glucose tolerance, were associated with an increased risk of stroke . Huang et al. performed a meta-analysis of prospective cohort studies to evaluate the associations between pre-diabetes, defined as impaired glucose tolerance, impaired fasting glucose, or raised HbA1c, and the risk of cardiovascular disease and all-cause mortality. Their analysis included 53 prospective cohort studies with 1,611,339 individuals. The median follow-up duration was 9.5 years. Compared with normo-glycaemia, pre-diabetes was associated with an increased risk of composite cardiovascular disease, stroke, and all-cause mortality. The health risk was increased in people with fasting glucose concentrations as low as 5.6 mmol/L or HbA1c of 39 mmol/mol. Increases in HBA1c to 39–47 was associated with an increased risk of composite cardiovascular disease and coronary heart disease. Follow-up of women who had GDM enables preventive measures and early diagnosis; early detection of DM decreases the risk of complications.

Gestational diabetes mellitus (GDM) is an independent risk factor for future type 2 diabetes mellitus, metabolic syndrome, cardiovascular morbidity, vascular endothelial dysfunction, renal and ophthalmic disease. The risk of these conditions may be decreased with proper prevention and interventions. However, the majority of women diagnosed with GDM do not undergo evaluation after pregnancy. According to one study, only 19% (4486 of 23,999) women who had GDM underwent testing to rule out type 2 DM within six months following pregnancy The authors of another study reported that women who developed GDM rarely followed the recommended dietary and physical activity guidelines in the postpartum period .This is despite the evidence that among women who had GDM, a modest post-partum weight loss resulted in significantly lower increases in fasting glucose and a significant reduction in 2-h glucose, while a 1-kg increase in weight was significantly associated with increase in fasting and 2-h glucose. Ferrara et al. reported that a lifestyle intervention that started during pregnancy and continued postpartum was feasible, prevented pregnancy weight retention, and helped overweight women lose weight .These findings and similar studies were the reason that Gabbe et al. published a “call for action”, an initiative of the National Diabetes Education Program, joined by the American College of Obstetricians and Gynecologists (ACOG) to promote opportunities for obstetrician-gynecologists (ob-gyns) and other primary care providers to better meet the long-term health needs of women with prior gestational diabetes mellitus (GDM), and their children .The authors stated that women with GDM face a lifelong increased risk for subsequent diabetes, primarily type 2 diabetes mellitus. Timely testing for pre-diabetes may provide an opportunity for ob-gyns to prevent or delay the onset of type 2 diabetes mellitus through diet, physical activity, weight management, and pharmacologic intervention. ACOG and the American Diabetes Association recommend testing women with a history of GDM at 6–12 weeks postpartum. If the postpartum test is normal, they should be retested every three years and at the first prenatal visit in a subsequent pregnancy. If pre-diabetes is diagnosed, women should be tested annually. Because children of GDM pregnancies face an increased risk for obesity and type 2 diabetes mellitus, families need support to develop healthy eating and physical activity behaviors. [7]

### 2.5 Preeclampsia

Preeclampsia is a complication of pregnancy. With preeclampsia, you might have high blood pressure, high levels of protein in urine that indicate kidney damage (proteinuria), or other signs of organ damage. Preeclampsia usually begins after 20 weeks of pregnancy in women whose blood pressure had previously been in the standard range. [8]

Preeclampsia is one of the leading causes of maternal morbidity and mortality worldwide. It is a pregnancy-specific multi-organ syndrome that affects 2–8% of pregnancies. It is a condition of placental pathogenesis with acute onset of predominantly cardiovascular manifestations attributable to generalized vascular endothelial activation and vasospasm resulting in hypertension and multi-organ hypo-perfusion. It is defined as a new onset of a multisystem pregnancy-related disorder that includes hypertension and either proteinuria or end-organ dysfunction, identified after 20 weeks gestation. During the index pregnancy, preeclampsia causes multi-system damages due to microangiopathy that lead to maternal morbidity that may include cardiac and renal failure, liver damage, cerebrovascular bleeding, pulmonary edema, disseminated intravascular coagulopathy (DIC), placental ischemia, etc. The effects on the fetus include prematurity (due to indicated preterm deliveries), fetal growth impairment, and intrauterine fetal demise. After delivery, the disorder tends to resolve in the majority of women although some remain hypertensive. There is a significant risk of preeclampsia recurrence in future pregnancies. There is an increased lifetime risk of chronic hypertension, cardiovascular disease (CVD), and stroke in women who experienced preeclampsia during pregnancy. The risk is related to the severity of the hypertensive disorder during pregnancy and the gestational age at the time of onset. The terms “preterm” or “early-onset” preeclampsia are used to delineate the severity of the disease in relation to the need for iatrogenic delivery before 37 weeks or the time of the diagnosis at or before 34 weeks of gestational age. Early-onset preeclampsia is especially associated with poor placentation, fetal growth restriction, and worse long-term maternal cardiovascular outcomes than late-onset preeclampsia, whose pathogenesis is more related to predisposing cardiovascular or metabolic risks for endothelial dysfunction.

Similar to the long-term risks of maternal morbidity associated with other pregnancy complications, e.g., gestational diabetes, it is unknown whether preeclampsia was actually the cause of increased risk of morbidity in these women or merely identified women who had a-priori increased risk of CVD morbidity. Researchers who believe in the former theory coined the expression, “maternal placental syndrome” (MPS), a term that combines various pregnancy complications (e.g., hypertensive disorders, placental abruption, infarction of the placenta, etc.) that originated (in their opinion) from “diseased placental vessels”, often in women who had metabolic risk factors for CVD (including obesity, pre-pregnancy hypertension, diabetes mellitus, and dyslipidemia). Ray et al. conducted a population-based retrospective cohort study of 1.03 million women, of whom 75,380 (7%) were diagnosed with a maternal placental syndrome, who were free from cardiovascular disease before their first documented delivery, in order to assess the risk of premature vascular disease in women who had had a pregnancy affected by maternal placental syndrome .They defined the following as maternal placental syndromes: preeclampsia, gestational hypertension, placental abruption, and placental infarction. Their primary endpoint was a composite of cardiovascular disease, defined as hospital admission or revascularization for coronary artery, cerebrovascular, or peripheral artery disease at least 90 days after the delivery discharge date. The incidence of cardiovascular disease was 500 per million person-years in women who previously had a maternal placental syndrome compared with 200 per million in women who had not (adjusted hazard ratio [HR] 2.0, 95 CI 1.7–2.2). This risk was higher in the combined presence of a maternal placental syndrome and poor fetal growth (3.1, 2.2–4.5) or a maternal placental syndrome and intrauterine fetal death .

Respondents with knowledge of the current guidelines of the German Association of Obstetrics and Gynecology concerning follow up and risk management of preeclampsia (45.2%) were more often aware of the development of CVD and stroke, and counseled patients on self-blood-pressure measurement, meaning, and long-term-risks of PE and attached importance to family history of PE, compared to physicians with no knowledge of the guidelines. The authors concluded that although the majority of obstetrician-gynecologists were aware of higher CVD risk after preeclampsia, weaknesses existed in the follow-up care and counseling of these patients. These deficiencies would be amendable to directed educational activities to improve the implementation of current guidelines. [9]

### 2.6 Preterm Deliveries

Preterm birth is when a baby is born too early, before 37 weeks of pregnancy have been completed. In 2021, preterm birth affected about 1 of every 10 infants born in the United States. The preterm birth rate rose 4% in 2021, from 10.1% in 2020 to 10.5% in 2021. However, racial and ethnic differences in preterm birth rates remain. In 2021, the rate of preterm birth among African-American women (14.8%) were about 50 percent higher than the rate of preterm birth among white or Hispanic women (9.5% and 10.2% respectively). [10]

Women who delivered prematurely are also at increased risk of long-term cardiovascular disease (CVD) and additional morbidities. Kessous et al. compared the incidence of cardiovascular morbidity in a cohort of 47,908 women, 5992 of whom (12.5%) delivered prematurely (<37 weeks’ gestation), between 1988–1999 with follow-up until 2010 Women who delivered prematurely (PTD) had higher rates of simple and complex cardiovascular events and higher rates of total cardiovascular-related hospitalizations. A linear association was found between the number of previous PTD and future risk of cardiovascular hospitalizations (5.5% for ≥2 PTDs; 5.0% for 1 PTD vs. 3.5% in the control group; p < 0.001). The association remained significant for spontaneous vs. induced PTD and for early (<34 weeks) and late (34 weeks to 36 weeks six days’ gestation) PTD. In a Cox proportional hazards model adjusted for pregnancy confounders such as labor induction, diabetes mellitus, preeclampsia, and obesity, PTD was independently associated with cardiovascular hospitalizations (adjusted hazard ratio, 1.4; 95% confidence interval, 1.2–1.6). In 2014 Robbins et al. summarized 10 studies that assessed the association between having a history of PTB and subsequent CVD morbidity or death .Compared with women who had term deliveries, women with any history of PTB had increased risk of CVD morbidity (variously defined; adjusted hazard ratio [aHR] ranged from 1.2–2.9; 2 studies), ischemic heart disease (aHR, 1.3–2.1; 3 studies), stroke (aHR, 1.7; 1 study), and atherosclerosis (aHR, 4.1; 1 study). Four of five studies that examined death showed that women with a history of PTB have twice the risk of CVD death compared with women who had term births.

Similar to other obstetric complications, it is unknown whether the premature delivery is the cause of later-life maternal morbidity or if it is the result of common predisposing factors that cause premature deliveries in these women. Immune or vascular abnormalities may be such factors. It is known that disorders of deep placentation and pathologic transformation of the spiral arteries are present in a wide range of pregnancy complications. In 2011 Romero et al. reported that placental-bed biopsies in women with placental syndrome showed failure of physiologic transformation of the spiral arteries [10]. Specifically, the mean percentage of spiral arteries that failed physiologic transformation during early pregnancy was significantly higher in women who experienced preterm labor or preeclampsia compared with women with uncomplicated term pregnancies. They proposed that changes in the population and function of immunocytes at the maternal-fetal interface can be part of the mechanism of disease of obstetric disorders.

A history of preterm delivery identifies women who should be targeted for CVD screening and preventative therapies. Future research is needed to assess the potential impact of such preventive measures on the incidence of CVD in this population. [10]

### 2.7 Recurrence and long-term maternal health risks of hypertensive disorders of pregnancy: a population-based study

Alice B. Andersgaard, MD [11] did a study on long term maternal health risk. The purpose of that study was to investigate the recurrence risk of hypertensive disorders in subsequent pregnancies and to explore the associations among hypertensive disorders of pregnancy and maternal cardiovascular risk factor profile and the development of cardiovascular diseases later in life [11]. Their main findings were that the women with a history of hypertensive disorder of pregnancy have an unfavorable cardiovascular risk profile, a higher frequency of CVD and hypertension, and more extensive carotid atherosclerosis. The increased risk of CVD among women with a history of preeclampsia is well-described. They showed that these women have an unfavorable risk profile based on history, physical examination, blood test results, and carotid artery ultrasound scanning. Their findings indicate that preeclampsia alters the CVD risk profile either by altering the maternal physiologic condition or by unmasking preexisting vulnerability that is related to constitutional predisposition. In their study, the objectively assessed CVD risk profile was different among women who had preeclampsia, gestational hypertension, or gestational proteinuria that supports the hypothesis that differences in the clinical presentation might be due to different underlying mechanisms that may have different implications for later CVD.19 Women with a history of preeclampsia had double the risk of hypertension and coronary artery disease compared with control subjects. They had carotid plaques more often, had larger total carotid plaque area and thicker intima-media layer compared with control subjects. Family history of CVD was more common among these women, which suggested that familial risk may be associated with underlying genetic predisposition to vascular dysfunction or other related factors (such as familial food habits, life style).

They found out that Preeclampsia in the first pregnancy increased the risk of recurrence in later pregnancies (relative risk, 6.6; 95% confidence interval, 5.5–7.9) compared with a normotensive first pregnancy. Women with a history of preeclampsia or nonproteinuric hypertension had an unfavorable cardiovascular risk profile. Hypertension was prevalent in 25% and 28% of the women, respectively. The carotid artery intima media thickness and total carotid plaque area were significantly larger in women with previous preeclampsia. A strong association between hypertensive disorders of pregnancy and an increased risk of atherosclerosis and cardiovascular diseases was demonstrated by the assessment of risk factors that can be potentially modified.

### 2.8 The maternal health clinic: an initiative for cardiovascular risk identification in women with pregnancy-related complications

Maria C. Cusimano, [12] did research an published a paper on an initiative for cardiovascular risk identification in women with pregnancy-related complications. The objective was Women who develop certain common pregnancy complications have a greater chance of developing cardiovascular disease (CVD) later in life. However, most health care providers do not provide postpartum cardiovascular risk counselling or follow-up. The Maternal Health Clinic was established to address this gap in care. It targets women at increased risk of CVD to inspire lifestyle changes, encourage long-term follow-up, and initiate primary prevention. they summarized results from the first 17 months of completed clinic visits.

They designed their study on patients experiencing at least one relevant complication in their index pregnancy were referred to the Maternal Health Clinic through standard postpartum order sheets. Patients underwent a complete assessment including screening history, physical examination, fasting bloodwork, and urinalysis. Lifetime and 30-year CVD risk estimates, along with a metabolic syndrome calculation, were determined for each patient.

The result of the study was Complications most commonly leading to referral were gestational diabetes or impaired glucose tolerance (32.7%), preeclampsia (29.3%), preterm birth (29.3%), and gestational hypertension (19.6%). The clinic analysis group (n = 92) was compared with a healthy control group from the Preeclampsia New Emerging Team study (n = 118). Patients in the clinic analysis group had significantly increased lifetime and 30-year CVD risk estimates compared with healthy controls (P < .0001). Furthermore, 17.4% of the clinic analysis group had metabolic syndrome, compared with 6.78% of healthy controls (P < .05).

Patients in the Maternal Health Clinic analysis group (n ¼ 92) were compared with a normotensive postpartum control group (n ¼ 118) from the Preeclampsia New Emerging Team (PE-NET) prospective longitudinal cohort, previously described.4,13,14 Normotensive women with a prior history of hypertension, diabetes, renal disease, or cardiovascular disease were excluded from the PE-NET cohort. Descriptive statistics were used to analyze the distribution of baseline characteristics in the Maternal Health Clinic analysis group and the PE-NET control group. Risk of CVD was determined for individual patients by 3 methods: a lifetime CVD risk estimate,19 a 30-year CVD risk estimate,18 and metabolic syndrome calculation.13,20 These measures were chosen because they help to better communicate risk to postpartum women who may have low short-term risk despite having significant risk factors for future CVD. The multiple comparison c Q1 3 test was used to compare the percentage of individuals in the Maternal Health Clinic analysis group and PE-NET control group who had optimal, non-optimal, elevated, or major lifetime risk estimates. The risk factors included in the lifetime CVD risk estimate included total cholesterol, systolic blood pressure, diastolic blood pressure, elevated fasting glucose or a previous diagnosis of diabetes, and smoking status. The Mann-Whitney U test was used to compare the mean 30-year risk estimate in each group. Finally, the c2 test was used to compare the prevalence of metabolic syndrome between groups. Ante hoc power analysis was not performed. Statistical analysis was performed with R version 2.13.2 and Q2 OpenEpi version 2.3.1.

The study demonstrated that the Maternal Health Clinic accurately identifies postpartum patients that have underlying cardiovascular risks which make them susceptible to CVD. The clinic may serve as an effective primary prevention strategy.

### 2.9: Maternal Healthcare in Migrant: A Systematic Review

Lígia Moreira Almeida [13] did a study in 2013 on Maternal Healthcare in Migrant. In the study the found out that Pregnancy is a period of increased vulnerability for migrant women, and access to healthcare, use and quality of care provided during this period are important aspects to characterize the support provided to this population. A systematic review of the scientific literature contained in the MEDLINE and SCOPUS databases was carried out, searching for population-based studies published between 1990 and 2012 and reporting on maternal healthcare in immigrant populations. A total of 854 articles were retrieved and 30 publications met the inclusion criteria, being included in the final evaluation.

The majority of studies point to a higher health risk profile in immigrants, with an increased incidence of co-morbidity in some populations, reduced access to health facilities particularly in illegal immigrants, poor communication between women and caregivers, a lower rate of obstetrical interventions, a higher incidence of stillbirth and early neonatal death, an increased risk of maternal death, and a higher incidence of postpartum depression. Incidences vary widely among different population groups. Some migrant populations are at a higher risk of serious complications during pregnancy, for reasons that include reduced access and use of healthcare facilities, as well as less optimal care, resulting in a higher incidence of adverse outcomes. Tackling these problems and achieving equality of care for all is a challenging aim for public healthcare services.

The most recent waves of immigration show the increasing feminization of this phenomenon. Migrant women frequently initiate the mobility process at a childbearing age, irrespective of individual motivations for leaving their countries. They are also frequently exposed to biological and psychosocial risks when confronted with new contexts, environments and lifestyles that tend to accentuate situations of social vulnerability. In addition to the anxieties inherent to migration, there is scientific evidence suggesting increased vulnerability during pregnancy and the postpartum period . Stressors associated with the migration process may be particularly important after delivery, exacerbating psychopathological complications such as the postpartum blues, psychosis and depression . Several psychosocial risk factors have been described, and include lack of social support, recent life stresses (including factors leading to migration and the migration process itself), personality variables and feelings about pregnancy or parenthood. Social and physical environmental adversity have been associated with maternal stress and pregnancy and infant health outcomes including prematurity, low-birth weight and infant mortality . Migrants also exhibit a greater risk of suffering from mental illness, including depression, schizophrenia and post-traumatic stress, as a result of specific psychosocial determinants . The issue of vulnerability acquires more alarming contours when there are barriers that hinder the access of migrant populations to health systems, such as those related to economic difficulties, language problems, issues, legal status, healthcare provider’s attitudes, and cultural differences . It is therefore important to assess the determinants of maternal healthcare in immigrant populations in order to establish policies that better attend these women’s requirements.

# Chapter 3

## Literature Review

### 3.1 Introduction

In this chapter some of the important attributes that we have used for developing our own predictor are described. We used 6 dependent variables and 1 independent variable metrics in our works and illustrates couple of mostly selected metrics/features in this chapter.

### 3.2 Feature Details

#### BodyTemp:

BodyTemp refers to body temperature in this dataset. Body temperature is one of the key features when a woman is pregnant. Medical research shows that overheating during pregnancy can put a baby at risk. Health guidelines advise that getting a mothers core body temperature at or over 102°F (39°C) can be too hot for the baby.

But also, it’s normal to feel somewhat warmer when a woman is pregnant. Several body changes during pregnancy can slightly raise mothers body temperature, and that’s completely fine. It’s when she is exposed to too much heat that she can feel unwell and it can affect how her baby develops. [14]

Body temperature plays a crucial role in maternal health because it serves as an indicator of various physiological processes and can provide insights into the well-being of both the mother and the developing fetus. Maintaining a stable body temperature is essential for normal fetal development and maternal health. Fluctuations in body temperature can be a sign of underlying health issues or complications that require medical attention. [15]

During pregnancy, several factors can affect body temperature regulation in expectant mothers. The metabolic rate increases, and hormonal changes occur, leading to changes in thermoregulation. The body's core temperature tends to rise slightly due to an increase in blood flow, hormonal changes, and an elevation in basal metabolic rate. However, the body's ability to dissipate excess heat may be impaired during pregnancy, making pregnant women more susceptible to overheating. Abnormal body temperature in pregnant women can indicate various conditions that may pose risks to both the mother and the fetus. For instance, a high body temperature, known as hyperthermia, can result from infections such as urinary tract infections, respiratory infections, or viral illnesses. Hyperthermia during pregnancy has been associated with an increased risk of miscarriage, birth defects, preterm labor, and developmental abnormalities in the fetus. It is crucial to promptly identify and manage infections to prevent potential complications. On the other hand, a consistently low body temperature, known as hypothermia, can also be problematic. Hypothermia may occur due to exposure to cold environments, inadequate clothing, or certain medical conditions. Prolonged or severe hypothermia can lead to poor fetal growth, preterm birth, and increased risks of complications during labor and delivery. [15]

Monitoring and maintaining a stable body temperature during pregnancy is essential for ensuring optimal maternal and fetal health. Healthcare providers regularly assess body temperature as part of routine prenatal care to identify any deviations from the normal range. Prompt medical intervention can be initiated if abnormal body temperature is detected. It is important to note that while body temperature is an important indicator of maternal health, it should be considered in conjunction with other clinical signs, symptoms, and laboratory investigations for accurate diagnosis and management.

#### HeartRate:

HeartRate refers to heart rate of the woman. During pregnancy, many changes happen that affect your entire body, including your heart and blood vessels. Over the course of pregnancy, your blood volume increases by almost 50%. This means your heart has to work harder to pump blood through your body. It sends much of this blood to your growing fetus. Your heart rate speeds up to get the job done. [16]

Heart rate is an essential parameter to monitor in maternal health because it provides valuable information about the cardiovascular system's functioning during pregnancy. The maternal cardiovascular system undergoes significant changes to support the growing fetus, and monitoring heart rate helps assess the overall maternal well-being. Here are the reasons why heart rate matters in maternal health:

* **Cardiac Output:** During pregnancy, a woman's cardiac output (the volume of blood pumped by the heart per minute) increases significantly to meet the demands of the developing fetus. Heart rate plays a crucial role in determining cardiac output since it represents the number of times the heart contracts in a minute. Monitoring heart rate helps assess the adequacy of cardiac output and the heart's ability to meet the increased demands of pregnancy. [17]
* **Maternal Hemodynamics:** Pregnancy induces changes in the maternal vascular system to ensure sufficient blood supply to the fetus. Heart rate reflects the maternal hemodynamic status, including the balance between heart rate, blood volume, and vascular resistance. Abnormal heart rates may indicate issues such as inadequate blood flow to the uterus and placenta, which can affect fetal growth and development. [18]
* **Early Detection of Complications:** Certain pregnancy complications can affect cardiovascular health, such as preeclampsia, gestational diabetes, or maternal heart conditions. Monitoring heart rate allows healthcare providers to identify abnormal rhythms, excessively high or low heart rates, or arrhythmias that may signify an underlying problem. Early detection enables timely intervention and management of these complications, reducing the risk of adverse maternal and fetal outcomes. [18]
* **Exercise and Physical Activity:** Regular physical activity during pregnancy is generally beneficial for maternal and fetal health. Monitoring heart rate during exercise helps women stay within safe limits, ensuring that the cardiovascular system does not become excessively strained. By maintaining an appropriate heart rate range, women can enjoy the benefits of exercise while minimizing potential risks. [19]
* **Anesthesia and Delivery:** Heart rate monitoring is essential during labor and delivery, particularly if interventions like anesthesia or cesarean section are required. Changes in heart rate patterns can indicate fetal distress or maternal complications that may necessitate immediate medical attention.

#### Age:

Age is an important factor in maternal health risk due to the physiological changes that occur in a woman's body as she ages, which can increase the likelihood of complications during pregnancy and childbirth. While pregnancy is generally considered a natural process, it places significant demands on a woman's body, and certain age groups may face additional challenges.

Research has shown that both younger and older maternal ages are associated with increased health risks. Teenage mothers (those under 20 years of age) tend to have higher rates of preterm birth, low birth weight, and pregnancy-related complications compared to women in their 20s and 30s. The reasons behind these risks include inadequate prenatal care, socio-economic factors, and biological immaturity of the teenage body for pregnancy and childbirth. [20]

On the other hand, advanced maternal age, typically defined as 35 years or older, is associated with its own set of risks. Women in this age group are more likely to experience fertility issues, chromosomal abnormalities in the fetus (such as Down syndrome), gestational diabetes, hypertension, and complications during labor. These risks increase with further age progression, particularly after the age of 40. [21]

The American College of Obstetricians and Gynecologists (ACOG) acknowledges the influence of maternal age on pregnancy outcomes and provides guidelines for healthcare providers to address the unique risks associated with different age groups. These guidelines emphasize the importance of preconception counseling, early prenatal care, and close monitoring of older pregnant women to mitigate potential complications.

#### BS:

Bs refers to blood sugar in this dataset. Blood sugar plays a vital role in maternal health.

High blood glucose levels during pregnancy can also increase the chance that your baby will be born too early, weigh too much, or have breathing problems or low blood glucose right after birth. High blood glucose also can increase the chance that you will have a miscarriage link or a stillborn baby [22]

Here are some reasons why blood sugar matters in maternal health:

• **Gestational diabetes** : Gestational diabetes is diabetes diagnosed for the first-time during pregnancy (gestation). Like other types of diabetes, gestational diabetes affects how to cells use sugar (glucose). Gestational diabetes causes high blood sugar that can affect pregnancy and baby's health. Gestational diabetes mellitus (GDM), a type of diabetes, manifests during pregnancy. Its distinguishing characteristic is high blood sugar levels, which often return to normal after birth. Between 2 and 10% of pregnancies worldwide are affected with GDM, which can increase the risk of complications for both the mother and the unborn child. Women with GDM are more likely to develop type 2 diabetes, preeclampsia, and cesarean sections later in life. Untreated GDM increases a mother's chance of having babies who are too big for their bodies, have respiratory issues, hypoglycemia, and other issues. [22]

• **Macrosomia**: Macrosomia refers to a baby who is considerably larger than normal. All of the nutrients the fetus receives come directly from the mother's blood. If the maternal blood has too much glucose, the pancreas of the fetus senses the high glucose levels and produces more insulin in an attempt to use this glucose. The fetus converts the extra glucose to fat. Even when the mother has gestational diabetes, the fetus is able to produce all the insulin it needs. The combination of high blood glucose levels from the mother and high insulin levels in the fetus results in large deposits of fat which causes the fetus to grow excessively large. [23]

• **Hypoglycemia**: Hypoglycemia refers to low blood sugar in the baby immediately after delivery. This problem occurs if the mother's blood sugar levels have been consistently high, causing the fetus to have a high level of insulin in its circulation. After delivery, the baby continues to have a high insulin level, but it no longer has the high level of sugar from its mother, resulting in the newborn's blood sugar level becoming very low. The baby's blood sugar level is checked after birth, and if the level is too low, it may be necessary to give the baby glucose intravenously. [23]

• **Preexisting diabetes**

Preexisting diabetes (also called pregestational diabetes) means have diabetes before get pregnant. This is different from gestational diabetes, which is a kind of diabetes that some women get during pregnancy. Women with diabetes can and do have healthy pregnancies and healthy babies. But untreated diabetes can cause complications for both moms and babies.

In the United States, about 1 to 2 percent of pregnant women have preexisting diabetes. The number of women with diabetes during pregnancy has increased in recent years. [24]

#### SystolicBP

Hypertension in pregnancy has traditionally been defined as a systolic blood pressure (sBP) of at least 140 mm Hg or a diastolic blood pressure (dBP) of at least 90 mm Hg, or both.1 Hypertension defined in this way identifies pregnant women at increased risk of pre-eclampsia and other maternal and fetal or neonatal complications, including death, and these women are recommended to receive enhanced antenatal care and monitoring worldwide. [22]

Blood pressure is the amount of pressure the blood places against the blood vessels walls with each heartbeat. A person can experience elevated or high blood pressure, or hypertension, during pregnancy. This is when the blood puts more pressure than normal against the artery walls.

According to the AHA Trusted Source, people living with untreated high blood pressure are at a greater risk of developing a heart attack, stroke, or other health issues, such as kidney disease.

During pregnancy, high blood pressure is also known as gestational hypertension.. Gestational hypertension occurs if the woman’s blood pressure is within the normal range for the first 20 weeks of pregnancy and then increases to 140/90 mm Hg or higher during the second half of the pregnancy. [23]

Preeclampsia is a complication of pregnancy. With preeclampsia, you might have high blood pressure, high levels of protein in urine that indicate kidney damage (proteinuria), or other signs of organ damage. Preeclampsia usually begins after 20 weeks of pregnancy in women whose blood pressure had previously been in the standard range. [24]

Preeclampsia is one high blood pressure (hypertension) disorder that can occur during pregnancy. Some maternal health conditions associated with elevated SBP during pregnancy include:

**• Gestational hypertension** is high blood pressure that begins after 20 weeks without problems in the kidneys or other organs. Some women with gestational hypertension may develop preeclampsia.

**• Chronic hypertension** is high blood pressure that was present before pregnancy or that occurs before 20 weeks of pregnancy. High blood pressure that continues more than three months after a pregnancy also is called chronic hypertension.

**• Chronic hypertension** with superimposed preeclampsia occurs in women diagnosed with chronic high blood pressure before pregnancy, who then develop worsening high blood pressure and protein in the urine or other health complications during pregnancy.

**• Eclampsia** is a severe form of preeclampsia that involves seizures. It poses significant risks to the health and well-being of both the mother and the baby and requires immediate medical attention. [24]

#### DaistolicBP:

Diastolic blood pressure is the pressure on the blood vessels when the heart muscle relaxes. The diastolic pressure is always lower than the systolic pressure.

* Diastolic blood pressure (DBP) is an important indicator of maternal health during pregnancy. It represents the pressure in the arteries when the heart is at rest between contractions. Maintaining a healthy diastolic blood pressure is crucial for several reasons.
* **Pre-eclampsia and Gestational Hypertension:** High diastolic blood pressure is one of the key criteria for diagnosing pre-eclampsia, a serious condition that affects pregnant women and can lead to complications such as organ damage and impaired fetal growth. Gestational hypertension, which is characterized by high blood pressure during pregnancy without the presence of proteinuria (protein in the urine), is also closely linked to diastolic blood pressure.
* **Fetal Health:** High diastolic blood pressure can impact the blood flow to the placenta, compromising the oxygen and nutrient supply to the developing fetus. This can result in fetal growth restriction and even stillbirth. Additionally, maternal hypertension increases the risk of preterm birth, which is associated with a higher likelihood of neonatal complications.
* **Maternal Organ Function:** Elevated diastolic blood pressure during pregnancy can strain the mother's organs, particularly the kidneys. It can lead to reduced renal blood flow and impair the kidneys' ability to filter waste products, potentially causing kidney dysfunction or damage.
* **Cardiovascular Risk:** Women with a history of high diastolic blood pressure during pregnancy have an increased long-term risk of developing hypertension and cardiovascular diseases later in life. [25]

Therefore, monitoring and managing diastolic blood pressure during pregnancy is crucial for identifying women at risk and implementing appropriate interventions to reduce future cardiovascular risks.

# Chapter 4

## Methodology

### 4.1 Introduction:

In this chapter the methodology that are used (Naïve Bayes, J48, Multilayer Perceptron, CHIRP) in Maternal health risk prediction and the architecture are briefly described.

### 4.2 Base Algorithm

In this paper, several base algorithms are used on the dataset. Those are briefly described below

#### 4.2.1 Naïve Bayes

Bayesian classifier assign the most likely class to a given example described by its feature vector [29]. Learning such classifiers can be greatly simplified by assuming that features are independent given class, that is, P(X|C) = (𝑋𝑖 | 𝐶), when X = (𝑋1, … … … , 𝑋𝑛 ) is a feature vector and C is a class. Despite this unrealistic assumption, the resulting classifier known as naïve Bayes is remarkably successful in practice, often competing with much more sophisticated techniques. Naïve Bayes has proven effective in many practical applications, including text classification, medical diagnosis, and systems performance management.

The success of naive Bayes in the presence of feature dependencies can be explained as follows: optimality in terms of zero-one loss (classification error) is not necessarily related to the quality of the fit to a probability distribution (i.e., the appropriateness of the independence assumption). Rather, an optimal classifier is obtained as long as both the actual and estimated distributions agree on the most-probable class. For example, prove naive Bayes optimality for some problems classes that have a high degree of feature dependencies, such as disjunctive and conjunctive concepts.

However, this explanation is too general and therefore not very informative. Ultimately, we would like to understand the data characteristics which affect the performance of naive Bayes. While most of the work on naive Bayes compares its performance to other classifiers on particular benchmark problems (e.g., UCI benchmarks), our approach uses Monte Carlo simulations that allow a more systematic study of classification accuracy on parametric families of randomly generated problems. Also, our current analysis is focused only on the bias of naive Bayes classifier, not on its variance. Namely, we assume an infinite amount of data (i.e., a perfect knowledge of data distribution) which allows us to separate the approximation error (bias) of naive Bayes from the error induced by training sample set size (variance).

We analyze the impact of the distribution entropy on the classification error, showing that certain almost deterministic, or low-entropy, dependencies yield good performance of naive Bayes. (Note that almost-deterministic dependencies are a common characteristic in many practical problem domains, such as, for example, computer system management and error correcting codes.) We show that the error of naive Bayes vanishes as the entropy H(P(X|C)) approaches zero. Another class of almost-deterministic dependencies generalizes functional dependencies between the features. Particularly, we show that naive Bayes works best in two cases: completely independent features (as expected) and functionally dependent features (which is less obvious), while reaching its worst performance between these extremes.

We also show that, surprisingly, the accuracy of naive Bayes is not directly correlated with the degree of feature dependencies measured as the class-conditional mutual information between the features, I(𝑋𝑖; 𝑋𝑗 | 𝐶) (𝑋𝑖 and 𝑋𝑖 are features and C is the class). Instead, our experiments reveal that a better predictor of naive Bayes accuracy can be 4 the loss of information that features contain about the class when assuming naive Bayes model, namely I(Instead, our experiments reveal that a better predictor of naive Bayes accuracy can be 4 the loss of information that features contain about the class when assuming naive Bayes model, namely I(C ; (𝑋𝑖 , 𝑋𝑗 )

– 𝐼𝑁𝐵 (𝐶 ; (𝑋𝑖 , 𝑋𝑗)),

where 𝐼𝑁𝐵 is the is the mutual information between features and class under naive Bayes assumption. [29]

#### 4.2.2 J48

C4.5/J48 is a widely used machine learning algorithm, which is a decision tree algorithm. [30]

This is a type of the ID3 algorithm, developed by Quinlan and is described in Fig. 1. The C4.5/J48 algorithm differs from the IDE3 as while building a decision tree, the algorithm can accept the continuous and the categorical attributes. Because of a high noise or a very detailed training data set, the J48 algorithm uses an enhanced technique of tree pruning for decreasing the misclassification error. Furthermore, this algorithm also used a greedy divide-and-conquer method for recursively inducing decision trees containing the database/dataset attributes for further classification. In any decision tree, classification is a major performance parameter. The classification error can be defined as the percentage of the misclassified cases. The C4.5 algorithm is seen to accept the continuous and the categorical attributes while developing a decision tree. This decision tree can be developed by making use of the top down or the bottom-up approach. Furthermore, the J48 classifier algorithm is divided into a dataset based on the different attribute values of the present data for separating a probable prediction. The decision tree contains many decision nodes and leaf nodes, wherein the decision nodes determine the test of the attributes while the leaf nodes represent the class values. Every path in the decision trees from the root to the leaf node determines the rule. This J48 classifier algorithm can develop its decision tree depending on the information of the theoretical attribute values of the present training data. Also, in the case of a J48 algorithm, every feature or attribute separately estimates the gain value and the calculation process is continued till the prediction process is completed. An appropriate feature is defined as the feature which gives a lot of information regarding the data instances. This feature can be classified as a root node if it consists of the maximal information gain. After selecting the root node, the J48 algorithm can divide the training data into many subsets which correspond to the various values of a chosen feature and this process is repeated for every subset till every subset is assigned to one class.

The J48 algorithm consists of many features described below:

• It is accessible as an open source in the WEKA interface in Java.

• The algorithm helps in building easy to understand models.

• The algorithm makes use of the categorical and the continuous values.

• The algorithm provides a technique known as the imputation, which deals with missing values. This technique helps in resolving the missing value problem, which is a significant feature, after determining the missing values based on the available data.

• The algorithm also provides a tree pruning process, which helps in building small trees and avoiding over-fitting of the data.

• Also, the algorithm provides the subtree replacement process which decreases the classification error after replacing the subtree with a leaf [30] .

#### 4.2.3 Multilayer Perceptron

A multilayer perceptron is a class of neural network that is made up of at least 3 nodes [31]. So now you can see the difference. Also, each of the node of the multilayer perceptron, except the input node is a neuron that uses a non-linear activation function.

The nodes of the multilayer perceptron are arranged in layers.

• The input layers

• The output layers

• Hidden layer: layers between the input and the output.

Also note the learning algorithm for the multilayer perceptron is known as black propagation (explained here).

**How the multilayer perceptron works**

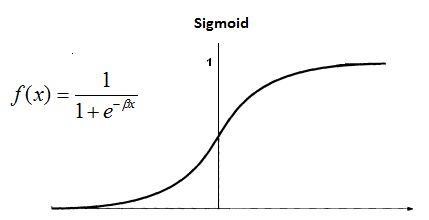
In MLP, the neurons use non-linear activation function that is designed to model the behavior of the neurons in the human brain.

A multi-layer perceptron has a linear activation function in all its neurons and uses backpropagation for its training.

**The activation functions**

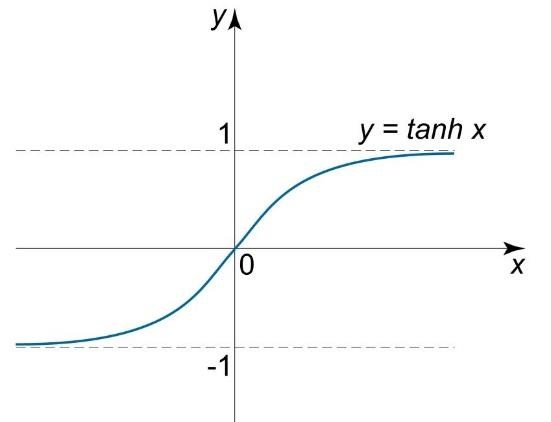
The activation function combines the input to the neurons with the weights and then adds a bias to produce the output. In other words, the activation function maps the weighted inputs to the output of the neuron.

One of such activation functions is the sigmoid function which is used to determine the output of the neuron. An example of a signed function is the logistic function which is shown below



**Figure 4.1:** Logistic function

Another example of a sigmoid function, is the hyperbolic tangent activation function shown below which produces an output ranging between -1 and 1



**Figure 4.2:** Hyperbolic tangent function

Yet another type of activation function that can be used is the Rectified Linear Unit or ReLU which is said to have better performance than the logistic function and the hyperbolic tangent function.

**Applying activation function to MLP**

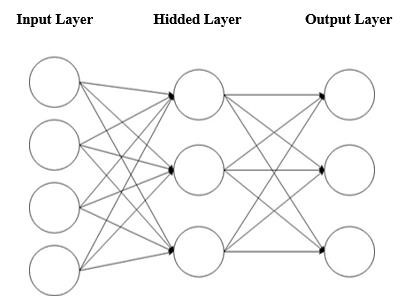
With activation function, we can calculate the output of any neuron in the MLP. Assuming w denotes the vector of weights, x is the vector of inputs, b is the bias and ϕ is the activation function, the for the ith, neuron, the output y would be given by:

y = ( ∑𝑛𝑖=1 𝑤𝑖𝑥𝑖 + b) = 𝜑(𝑤𝑇𝑥 + b) ……………………… (1)

An MLP is made up of a set of nodes which forms the input layer, one or more hidden layers, and an output layer.

**Layers of Multilayer Perceptron (Hidden Layer)**

Remember that from the definition of multilayer perceptron, there must be one or more hidden layers. This means that in general, the layers of an MLP should be a minimum of three layers, since we have also the input and the output layer. This is illustrated in the figure below.



**Figure 4.3:** Multilayer Perceptron

Also, to note is that the function activating these hidden layers has to be non-linear function (activation function) as discussed in previous section.

**Training/Learning in Multilayer Perceptron**

The training process of the MLP occurs by continuous adjustment of the weights of the connections after each processing. This adjustment is based on the error in output (which is the different between the expected result and the output). This continuous adjustment of the weights is a supervised learning process called ‘backpropagation’.

The backpropagation algorithm consists of two parts:

* + Forward pass
  + Backward pass

In the forward pass, the expect output corresponding to the given inputs are evaluated. In the backward pass, partial derivatives of the cost function with respect to the different parameters are propagated back through the network. The process continues until the error is at the lowest value [31].

### 4.3 Proposed Algorithm

In this paper, a proposed algorithm is used on the dataset. It is briefly described below.

#### 4.3.1 CHIRP (Composite Hypercubes on an Iterated Random Projections)

CHIRP (Classifier based on Composite Hypercubes on Iterated Random Projections) is a classification algorithm that combines the concepts of composite hypercubes and iterated random projections. It aims to provide an efficient and accurate method for high-dimensional data classification. [32]

The algorithm starts by applying the Iterated Random Projections (IRP) technique to reduce the dimensionality of the input data. IRP maps the high-dimensional data into a lower-dimensional space while preserving the structure and discriminative information of the original data.

After dimensionality reduction, CHIRP constructs composite hypercubes, which are multidimensional structures that partition the reduced space into smaller regions. These hypercubes serve as decision boundaries for classification. Each composite hypercube represents a class, and the goal is to assign data points to the appropriate hypercube/class.

The construction of composite hypercubes involves dividing the reduced space into equal-sized subspaces and assigning data points to the corresponding subspaces based on their proximity. This process allows CHIRP to capture the local structures and relationships among data points in the reduced space.

During the classification phase, CHIRP assigns a test data point to a specific class by identifying the composite hypercube that contains the point in the reduced space. This assignment is based on the proximity of the data point to the hypercube boundaries.

The advantages of CHIRP include its ability to handle high-dimensional data, efficient computation, and the ability to capture complex data distributions. By combining the power of iterated random projections for dimensionality reduction and composite hypercubes for classification, CHIRP provides a robust and accurate classification algorithm.

Here is an overview of the CHIRP algorithm:

1. **Preprocessing**:
   * Input: Data matrix X of size N x D (N samples, D dimensions), where each row represents a data point and each column represents a feature.
   * Specify the desired target dimensionality d.
2. **Iterated Random Projections**:
   * Initialize a random projection matrix R of size D x d.
   * Project the data matrix X into the lower-dimensional space: Y = X \* R.
3. **Composite Hypercube Construction:**
   * Initialize an empty composite hypercube structure.
   * For each data point y in the reduced space Y:
     + Determine the corresponding hypercube in the composite hypercube structure based on the boundaries defined by its dimensions.
     + Update the representative point within the hypercube by aggregating the data point y.
4. **Classification:**
   * Given a new data point x, project it into the lower-dimensional space: y = x \* R.
   * Find the hypercube in the composite hypercube structure that contains the point y.
   * Assign the class label of the representative point within the identified hypercube to the new data point x.

#### 4.3.1 The CHIRP Algorithm

**Input:**

- Data matrix X of size N x D (N samples, D dimensions)

- Target dimensionality d (lower-dimensional space)

- Number of composite hypercubes k (number of classes)

**Procedure CHIRP(X, d, k):**

1. Apply the Iterated Random Projections (IRP) algorithm to reduce the dimensionality of the data:

reduced\_X = IRP(X, d)

2. Divide the reduced space into k equal-sized subspaces:

subspaces = divide\_space(reduced\_X, k)

3. Initialize composite hypercubes for each class:

hypercubes = []

for i = 1 to k:

hypercube = create\_hypercube(subspaces[i])

hypercubes.append(hypercube)

4. Assign data points to the composite hypercubes based on proximity:

for j = 1 to N:

point = reduced\_X[j]

class\_label = assign\_to\_hypercube(point, hypercubes)

assign\_class\_label(X[j], class\_label)

5. Return the assigned class labels for the data points X

**Output:**

- Class labels assigned to the data points X

CHIRP combines the efficiency of composite hypercubes for organizing and summarizing high-dimensional data with the dimensionality reduction capabilities of iterated random projections. By reducing the dimensionality and organizing the data points into hypercubes, CHIRP can efficiently classify new data points based on their location within the composite hypercube structure. [32]

### 4.4 System Architecture

The system architecture of the CHIRP algorithm typically involves several components working together to perform dimensionality reduction, composite hypercube construction, and classification. Here is a high-level overview of the system architecture:

1. **Data Preprocessing:** The input data is preprocessed to handle any missing values, normalize features, and handle other data-specific requirements. This step ensures that the data is in a suitable format for further processing.
2. **Iterated Random Projections (IRP):** The IRP technique is applied to reduce the dimensionality of the input data. This involves iteratively projecting the data onto lower-dimensional spaces using random projection matrices. The IRP component performs the necessary calculations and generates the reduced-dimensional representation of the data.
3. **Composite Hypercube Construction:** Once the data is projected into the lower-dimensional space, the composite hypercube construction component partitions the reduced space into composite hypercubes. The construction process involves dividing the space into equal-sized subspaces and assigning data points to the appropriate subspaces based on their proximity. This step creates the composite hypercubes that represent different classes.
4. **Classification:** In the classification phase, the system takes a new, unseen data point and assigns it to a specific class based on its proximity to the boundaries of the composite hypercubes. This process involves calculating the distance or similarity between the data point and the hypercube boundaries and making a decision based on predefined thresholds or rules.
5. **Training and Evaluation:** The CHIRP algorithm typically involves a training phase where the system learns the boundaries of the composite hypercubes and their association with different classes. This training phase uses labeled data to optimize the hypercube construction and classification parameters. Once trained, the system can be evaluated using testing data to assess its performance in terms of accuracy, precision, recall, and other evaluation metrics.
6. **Integration and Deployment:** The CHIRP algorithm can be integrated into larger software systems or frameworks for real-world applications. This may involve integrating with data pipelines, user interfaces, or other components to facilitate data input, visualization, and interaction with the algorithm. Deployment considerations include scalability, efficiency, and adaptability to handle large datasets and real-time classification requirements.