

https://dspace.org/

Detecting hearing loss in high-risk neonates using machine learning

Ismail, Safiyyah

2021
https://hdl.handle.net/10539/33857

Downloaded from a DSpace repository. https://dspace.org/

Detecting Hearing Loss in High-Risk Neonates Using Machine Learning

---

Miss Safiyyah Ismail

Supervisor(s):

Dr Selvarani Moodley

Professor Rod Alence



Masters Research Project in Data Science

submitted in partial fulfillments of the requirements for the degree of

Master of Arts in the field of eScience

in the

School of Social Science

University of the Witwatersrand, Johannesburg

31 March 2021

Declaration

I, Miss Safiyyah Ismail, declare that this research report is my own, unaided work. It is being submitted for the degree of Master of Arts in the field of e-Science at the University of the Witwatersrand, Johannesburg. It has not been submitted for any degree or examination at any other university.



Miss Safiyyah Ismail

31 March 2021

Abstract

Deafness is one of the most commonly occurring birth conditions in children worldwide creating an increasingly significant global health concern. Failure to early identify hearing loss and provide subsequent intervention services will likely have negative consequences on language, cognition, and socio-emotional development. Current approaches in detecting neonatal hearing loss are limited specifically in developing countries such as South Africa. Machine learning offers an opportunity to create models which could predict the likelihood of a hearing loss occurring in high-risk neonates allowing for early identification and intervention to occur. Thus, the main aim of the current study was to use predictive modelling to predict the likelihood of hearing loss in high-risk neonates. The study sample comprised of 12 044 male and female hearing and deaf and/or hard-of-hearing South African children who either formed part of the HI HOPES or universal newborn screening programme implemented at the Netcare Hospital Group. A nonexperimental, predictive modelling design was employed for the purpose of the current study. Predictive variables used in the current study included mode of delivery, prematurity, gestational age, family history of hearing loss, extracorporeal membrane oxygenation (ECMO), in-utero infections, craniofacial anomalies, physical findings, syndromes associated with hearing loss, neurodegenerative disorders, cultural-positive infections, meningitis, maternal and/or infant HIV infection, and ototoxic medication. The results from several Chi-Square (X^2) analyses showed significant correlations between each birth type (i.e., natural, elective caesarean, emergency caesarean), prematurity, family history, ECMO, in-utero infection, craniofacial anomalies, physical findings, syndromes associated with hearing loss, cultural-positive postnatal infections, meningitis, maternal and/or HIV infection, and ototoxic medication. The predictive models for hearing loss in high-risk neonates were developed using logistic regression and random forest (RF) classifiers. The major predictors of neonatal hearing loss determined by both models were prematurity, family

history, cultural-positive infections, and meningitis. The final reduction of error rate for the logistic regression was 90% with a prediction rate of 92%. In contrast, the random forest performed slightly poorer with an out-of-bag error rate of 14.8% and a prediction rate of 88%. The results of the current study demonstrated that machine learning algorithms can be used as potential tools for the evaluation and prediction of hearing loss in high-risk neonates.

Acknowledgements

I wish to express my deepest gratitude to my Divine Creator for opening this door of opportunity for me. I wish to express my sincere gratitude to my supervisors, Dr Selvarani Moodley and Professor Rod Alence for their constant guidance, expertise, and support. I would like to thank my family and friends for all their love and support.

This research project was funded by the DST-funded National e-Science Postgraduate Teaching and Training Platform (NEPTTP). Databases are provided by the Home Intervention Hearing (and) language Opportunities Parent Education Services (HI HOPES) Programme and the Netcare Hospital Group.

Table of Contents

Abstract iii

Acknowledgements iv

Table of Contents vi

List of Figures ix

List of Tables x

List of Abbreviations xi

Chapter 1: Introduction and Literature Review 7

1.1. Introduction 7

1.2. Newborn and Early Childhood Hearing Loss 7

1.3. Defining Childhood Deafness 9

1.4. Ecological View of Deafness 10

1.4.1. Bronfenbrenner's Bioecological Theory of Human Development 12

1.4.1.1. The Microsystem 12

1.4.1.2. The Mesosystem 15

1.4.1.3. The Ecosystem 16

1.4.1.4. The Macrosystem 17

1.5. Early Hearing Detection and Intervention (EHDI) and Universal Newborn Hearing Screening (UNHS) Programmes 18

1.5.1. Electrophysiological Testing in UNHS 19

1.5.1.1. Automated Otoacoustic Emissions 19

1.5.1.2. Automated Auditory Brainstem Response 20

1.5.2. EHDI and UNHS in South Africa 22

1.6. Machine Learning in Healthcare 24

1.6.1. Machine Learning in Audiology 26

1.6.2. Extending Quantitative Research in Neonatal and Early Childhood Audiology with Machine Learning 27

1.7. The Current Study 28

1.7.1. Study Aims 28

1.7.2. Study Objectives 28

1.7.3. Research Question 28

Chapter 2: Methodology 30

2.1. Dataset Description 30

2.1.1. Sampling Methods 30

2.1.2. Secondary Datasets: HI HOPES and Netcare Hospital Group 31

2.1.2.1. HI HOPES Dataset 31

2.1.2.2. Netcare Hospital Group Dataset 32

2.1.3. Patient Records 33

2.2. Research Design 34

2.2.1. Outcome and Predictor Variables 34

2.2.2. Classification: A Machine Learning Technique 35

2.2.2.1. Logistic Regression 36

2.2.2.2. Random Forest 37

2.3. Procedure 37

2.3.1. Handling Missing Values 38

2.3.2. Handling Class Imbalance 39

2.3.3. Statistical and Machine Learning Analyses 40

2.4. Ethical Considerations 41

Chapter 3: Results 43

3.1. Descriptive Analyses 43

3.1.1. Patient records: Neonate Profile 43

3.1.1.1. Mode of Delivery 43

3.1.1.2. Gestational Age 44

3.1.2. Risk Factors 45

3.1.3. Statistical Analyses 46

3.1.4. Classification Models 47

3.1.4.1. Logistic Regression 47

3.1.4.2. Random Forest 50

Chapter 4: Discussion 52

4.1. Contributions of the Current Study 58

4.2. Limitations and recommendations of the Current Study 59

Chapter 5: Conclusion 61

References 62

Appendices 85

Appendix A: Plagiarism Declaration 86

Appendix B: Ethical Approval 88

Appendix C: Supervisor-Supervisee Contract 90

Appendix D: Permission to Use Datasets 92

List of Figures

Figure 1:	Bronfenbrenner's Ecological Systems Model	13
Figure 2:	Variable Importance Plot of Predictors in Random Forest Risk Model	51

List of Tables

Table 1:	Frequencies and Percentages of HPCSA Defined Risk Factors Present in Sample	45
Table 2:	Pearson's Chi-Square (χ^2) Test of Contingencies Associations Between Risk Factors and Hearing Loss	46
Table 3:	Logistic Regression Coefficients for Gestational Age as a Predictor of Hearing Loss	47
Table 4:	Predictor Coefficients for the Logistic Regression Model Predicting Hearing Loss	49

List of Abbreviations

AABR	**A**utomated **A**uditory **B**rainstem **R**esponse
ABR	**A**uditory **B**rainstem **R**esponse
AIDS	**A**cquired **I**mmune **D**eficiency **S**yndrome
dB	**D**ecibel
EHDI	**E**arly **H**earing **D**etection (and) **I**ntervention
HI HOPES	**H**ome **I**ntervention **H**earing (and) language **O**pportunities **P**arent **E**ducation **S**ervices
HIV	**H**uman **I**mmunodeficiency **V**irus
HPCSA	**H**ealth **P**rofessions **C**ouncil of **S**outh **A**frica
JCIH	**J**oint **C**ommittee (on) **I**nfant **H**earing
kHz	**K**ilohertz
OAE	**O**toacoustic **E**missions
SKI-HI	**S**ensory **K**ids **I**mpaired **H**ome **I**ntervention
TB	**T**uberculosis
TNHS	**T**argeted **N**ewborn **H**earing **S**creening
UNHS	**U**niversal **N**ewborn **H**earing **S**creening
WHO	**W**orld **H**ealth **O**rganisation

Chapter 1: Introduction and Literature Review

1.1. Introduction

Hearing loss is one of the most increasingly significant global health concerns. It has long-term effects in the domains of cognition, language, psychosocial development, and brain organization (Kotby, Tawfik, Aziz, & Taha, 2008; Copley & Friderichs, 2010; Olusanya, Neumann, & Saunders, 2014). There is an overall and universal understanding that the first few years of an infant's life is crucial to laying down the foundation for optimal future development and growth (Storbeck & Moodley, 2010). Failure to early identify hearing loss at birth (or shortly thereafter) and provide subsequent intervention before the age of 6 months will likely have negative consequences on language, cognitive and socio-emotional development, as well as future scholastic achievement (Morgan & Vernon, 1994). Future effects are noted in the areas of vocational attainment (Bezuidenhout, 2016) and societal integration (Storbeck & Calvert-Evans, 2008). Thus, early identification of hearing loss can significantly minimize, if not avoid, the adverse consequences of congenital and early acquired hearing loss (Olusanya, 2008).

1.2. Newborn and Early Childhood Hearing Loss

Deafness has been reported as the most commonly occurring birth condition in children worldwide (Khan, Joseph, & Adhikari, 2018). Global estimates from the World Health Organisation (WHO) indicate that close to 466 million (5%) individuals worldwide are deaf or hard-of-hearing, with 34 million (7%) being children (WHO, 2020). Developing regions such as South Asia, Asia Pacific, and Sub-Saharan Africa have the highest prevalence of deafness (in varying degrees), with two-thirds of the world's deaf and hard-of-hearing population residing in these regions (Tucci, Merson, & Wilson, 2010; WHO, 2020). Sub-

Saharan Africa in particular is comprised of the underdeveloped countries in the world (McPherson & Swart, 1997). Due to the higher prevalence of environmental risk factors and scarcity in resources, children in Sub-Saharan Africa are more susceptible to pathologies related to childhood deafness (McPherson & Swart, 1997; Olusanya et al., 2014). These risk factors include: pre-, peri-, and postnatal complications; ototoxic medication; and infectious diseases such as meningitis, rubella, and measles (HPCSA, 2018). While the high prevalence of deafness in Sub-Saharan Africa may be explained by these risk factors, notable causes of deafness have not been well-documented (Kanji, 2016). Half of congenital and early-onset deafness is attributed to genetic causes with other contributing risk factors and causes likely varying between countries (Olusanya, 2011).

Although more than 50% of the risk factors associated with childhood deafness are preventable, resources in developing countries are often preferentially allocated to funding the prevention and intervention of life-threatening diseases such as the Human Immunodeficiency Virus (HIV) and Acquired Immune Deficiency Syndrome (AIDS), Tuberculosis (TB), and malaria (Abdalla & Omar, 2011). Despite the significant burden the effects of deafness pose on both the individual and society at large, it is not considered a priority. As a result, early identification, prevention, and intervention services for the deaf and hard-of-hearing population are largely neglected (Mackenzie & Smith, 2009; Abdalla & Omar, 2011). Furthermore, the criteria used to classify deafness in terms of laterality, frequency and intensity is not universal and may vary between countries which may account for the lack of urgency regarding the impact of childhood deafness (Berninger & Westling, 2011).

1.3. Defining Childhood Deafness

While there is much controversy surrounding the correct terminology to use when referring to hearing difficulties, for the purpose of the current study the term "hearing loss" will be used when referring to deaf and hard-of-hearing children regardless of whether it occurred congenitally or was progressive/acquired[1]. Furthermore, the term "hearing loss" as used in the current study is inclusive of children who are deaf (i.e., having a profound hearing loss) and hard-of-hearing (i.e., having a hearing loss ranging from mild to severe).

In South Africa, the minimum criterion for diagnosing hearing loss in children is a permanent hearing threshold of at least 40 decibels (dB) or greater averaging over frequencies of 0.5, 1.2, and 4 kilohertz (kHz) (HPCSA, 2018). This definition, however, excludes hard-of-hearing children with unilateral and/or milder forms of hearing loss (Bezuidenhout, 2016). While these children may experience developmental delays of a lesser degree when compared with those with a more severe hearing loss, they should still be considered important, especially for children with prelingual hearing loss. Early-onset hearing loss in children may occur congenitally or by manifesting postnatally as a late-onset, progressive, or acquired hearing loss (Olusanya, Luxon, & Wirz, 2005). Regardless of the aetiology, hearing loss occurs in varying degrees of severity ranging from mild to profound and affecting one (unilateral) or both (bilateral) ears (Nunez-Batalla, Jaudenes-Casaubon, Sequl-Canet, Vivanco-Allende, & Zubicaray-Ugarteche, 2016). Moreover, based on the locality of the hearing loss in the auditory system, hearing loss can be mixed, sensorineural, conductive, or auditory neuropathy (Kanji, 2016). Since it has been shown that all types, causes, and severities of hearing loss have an impact on development, there has been an international drive towards early identification and intervention services with the goal to

[1] The WHO uses "hearing loss" as an umbrella term to refer to individuals with hearing thresholds which deviate from normal. This term of "hearing loss" is used regardless of type, laterality, or degree of hearing loss, and whether the hearing loss is congenital or acquired (WH0, 2020)

sustain typical development outcomes for deaf and hard-of-hearing children (Moodley, 2016).

The development of deaf and hard-of-hearing children takes place in the context of complex systematic interactions between numerous factors within the physical, social, and cultural spheres of life (Kossewska, 2016). As such, understanding deafness through an ecological perspective aids in the identification of the factors within these biopsychosocial levels that influence the development of deaf and hard-of-hearing children and how they can be rectified to promote optimal and normative development (Harvey & Dym, 1987). An ecological view of deafness is presented below.

1.4. Ecological View of Deafness

Generational changes in children's status has led to an international recognition of the importance of children's rights (Melton, 2008), notably leading to the establishment of the United Nations Convention on the Rights of Children in 1989. Unfortunately, the principles set out in the Convention have not been fully and effectively implemented globally. Millions of children under the age of 5 years continue to suffer violations of their basic human rights by living in adverse environmental conditions (Grantham-McGregor et al., 2007). As a result, many children fail to reach their optimal development potential due to poverty, poor and unstimulating living conditions, inadequate sanitation, lack of adequate healthcare, malnutrition, and exposure to chronic and incurable diseases (Grantham-McGregor et al., 2007; Olusanya, 2011). These developmental barriers have detrimental long-term effects on the key interdependent domains of cognitive, sensorimotor, emotional, and psychosocial development (Johnson & Blasco, 1997; Grantham-McGregor et al., 2007; Olusanya, 2008; Olusanya, 2011).

Childrens' trajectories for physical, psychosocial, and cognitive development are set extremely early in life. The first 3 to 5 years of a child's life are vital as it marks the period in which brain development is highly and continuously influenced by the interaction between intrinsic (gene) and extrinsic (environmental) factors (Johnson & Blasco, 1997; Olusanya, 2011). These neurodevelopmental patters occur predictably and sequentially with each developmental skill building on the former (Johnson & Blasco, 1997). Minor delays in one or more stages of these processes can have long-term consequences on the structure and functionality of the developing brain (Johnson & Blasco, 1997). Fragile foundations for later developmental achievement may lead to a deficit in cognition, low educational enrollment and achievement, lack of vocational attainment, and ultimately, a substandard quality of life (Grantham-McGregor et al., 2007; Olusanya, 2011).

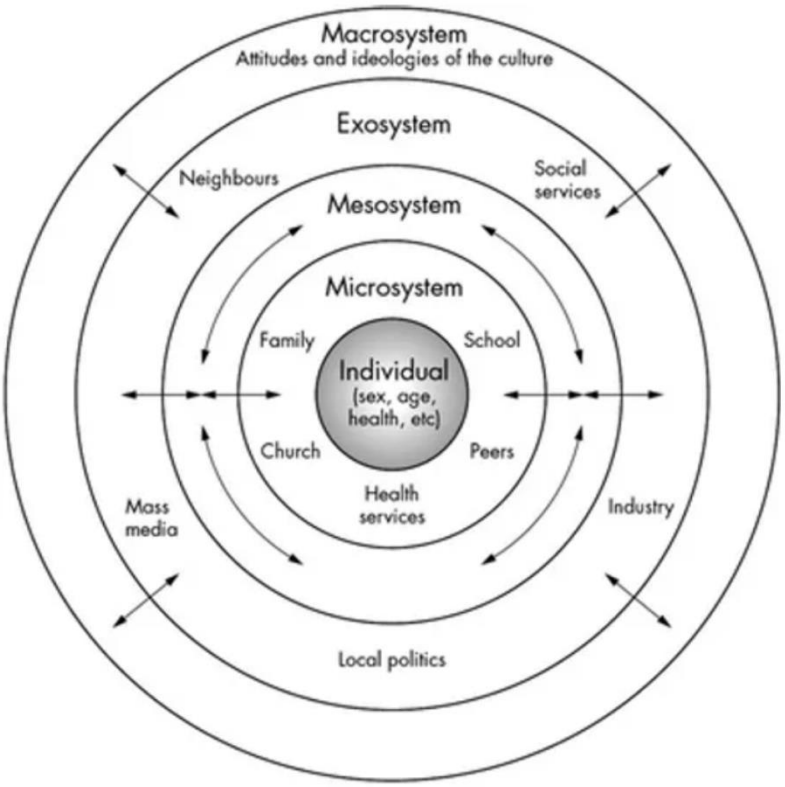
This is particularly apparent in sensory deficits such as hearing loss where it is considered that undetected hearing loss not only compromises optimal development in linguistic and communicative domains but in all other developmental domains as well (Olusanya, 2011; Storbeck & Moodley, 2011). As this is the case, both the JCHI and the HPCSA have provided guidelines and benchmarks for early detection and intervention services (HPCSA, 2018; JCHI, 2019). Regardless of these principles, the issue of early detection and intervention in South Africa is magnified by the lack of adequate healthcare services, parental lack of awareness, as well as individual perceptions and cultural beliefs (Olusanya, 2008). While deafness begins as a physiological condition, the way it affects an individual depends on how they effectively engages in their physical, social, and cultural environment (Harvey & Dym, 1987; Olusanya, 2008). Hence, it is therefore important to understand the experience and outcome of deafness on childhood development from an ecological perspective (Harvey & Dym, 1987).

1.4.1. Bronfenbrenner's Bioecological Theory of Human Development

Urie Bronfenbrenner's bioecological theory of human development is one of the most widely known theoretical frameworks of human development (Velez-Agosto, Soto-Crespo, Vizcarrondo-Oppenheimer, Vega-Molina, & Coll, 2017). Bronfenbrenner (1994) argued that for one to understand human development, one must consider the ecology wherein children exist and how it influences their growth and development (Stuart, 2009; Visser & Moleko, 2012). Bronfenbrenner (1979) describes this environment where development takes place as a set of systems fitting into each other like Russian Nesting Dolls. Higher levels contain the lower levels, and all levels are interrelated and interdependent on one another (Visser & Moleko, 2012). He proposed that factors influencing optimal development such as the aetiology of the illness, disability, or disorder; the application of effective intervention methods to deal with such difficulties; and promoting optimal health can only be understood if the ecological system of the child was considered (Stuart, 2009). This system is comprised of 4 subsystems that are related to and interact with one another (Bronfenbrenner, 1994) (see Figure 1). The synergy between each subsystem affects the entire system as a whole, which together supports and guides human growth and development (Visser & Moleko, 2012).

Figure 1

Bronfenbrenner's Ecological Systems Model



(Morris, 2010)

1.4.1.1. The Microsystem

The microsystem is the immediate environment wherein the child exists and which has the most influence on development (Bronfenbrenner, 1979). Interpersonal relationships, activities, and social roles experienced in the microsystem through face-to-face family, school, and peer group settings influence both the child's perception of their environment and in shaping future interactions (Bronfenbrenner, 1979; Bruyere & Garbarino, 2010). Most importantly perhaps, are the reciprocal interactions which occur between the child and their

caregiver (Bruyere & Garbarino, 2010). The purpose of these interactions generally leads to the formulation of an attachment relationship and subsequently, the construction of 'working models' or internal representations of the self and others (Ainsworth & Bowlby, 1991; Levy & Blatt, 1999; Bruyere & Garbarino, 2010). These internal mental representations lay the foundation for future healthy or unhealthy development to occur (Ainsworth & Bowlby, 1991). Internal and external factors which affect parenting/caregiver behaviour will ether pose as a risk or as an opportunity to optimal childhood development (Bruyere & Garbarino, 2010). Familial behavioural patterns, perceptions about deafness, emotional responses to deafness, and the overall interactions family members have with the deaf and hard-of-hearing child have powerful influences on their development (Harvey & Dym, 1987). Research has found that deaf and hard-of-hearing children are at increased risk to various forms of child maltreatment including physical, sexual, and emotional abuse and neglect (Kvam, 2008; Knutson, Johnson, & Sullivan, 2004; Schenkel, Rothman-Marshall, Schlehofer, Towne, Burnash, & Priddy, 2014). Barriers to adequate communication between parents and their deaf or hard-of-hearing child is likely to add strain to the parent-child relationship. This may add to increased levels of frustration and harsher physical discipline strategies which could subsequently lead to emotional and physical abuse (Schenkel et al., 2014). For example, mothers of children who are deaf or hard-of-hearing are more likely to use physical punishment in response to destructive behaviour than mothers of normal hearing children (Knutson et al., 2014). Additionally, their inability to communicate effectively make deaf and hard-of-hearing children easy targets for abuse and maltreatment as they are less likely to report instances of victimisation to parents or other adults (Schenkel et al., 2014). Many children who fall victim to maltreatment experience both short- and long-term socioemotional and physical difficulties (Cicchetti, Toth, & Maughan, 2000; Lansford, Dodge, Pettit, Bates, Crozier, & Kaplow, 2002; Olusanya, 2008). Poor parent-child

relationships have been found to be associated with greater mental health outcomes, poorer quality of life, negative self-perception, and re-victimisation amongst deaf and hard-of-hearing children (Burnash, Rothman-Marshall & Schenkel, 2010; Kushalnagar, Topolski, Schick, Edwards, Skalicky, & Patrick, 2011). Understanding the early trauma and victimisation of deaf and hard-of-hearing children can facilitate the implementation of more effective and tailored prevention and intervention services.

1.4.1.2. The Mesosystem

The mesosystem is comprised of the linkages between microsystems (Bronfenbrenner, 1994). Bronfenbrenner (1979) stressed that childhood development will be enhanced if the settings in which the child exists are strongly linked. The strength of these linkages depends on the reciprocal transactions between settings (Bronfenbrenner, 1994). Mothers who have actively engaged with medical health professionals through regular prenatal visits are able to form relationships with preofessionals who can provide them with vital information on their developing child both before and after birth (Bruyere & Garbarino, 2010). In South Africa, the majority of these consultations are conducted by nursing professionals at local clinics where pregnant and lactating women, children with disabilities, and children under the age of 6 can receive free access to healthcare services. It is within these settings where children who are deaf and hard-of-hearing (or who are at-risk for developing a hearing loss) can be early identified and receive referrals for appropriate intervention services (Moodley, 2012). Mothers who are made aware of their child's hearing status can assist their child's development by creating learning-rich environments and explore ways in which to communicate with their child. This awareness will stimulate a typical path of development for the developing child. Through this, parental frustration and stress may be minimized; and any form of child maltreatment may be avoided. Unfortunately, the

likelihood of mothers engaging effectively with healthcare professionals regarding the health and wellbeing of their children is limited due to the poor allocation of funding, shortage of medical staff, and lack of adequate healthcare facilities in South Africa (Moodley & Storbeck, 2012; Olusanya et al., 2014; Moodley & Storbeck, 2017).

1.4.1.3. The Exosystem

The exosystem consists of the interconnections between the micro- and mesosystems as well as those systems with which the developing child has no direct contact but which may indirectly influence the processes within their immediate setting (Bronfenbrenner, 1979; Bronfenbrenner, 1994). Exosystems are both informal and formal such as the parent's workplace, social networks, and neighbourhood characteristics (Algood, Hong, Gourdine, & Williams, 2011). The quality of the parent-child relationship can be highly influenced by both parenting stress and parental social support (Crouch & Behl, 2001). Parents who have little to no social support due to low socioeconomic status and unemployment are more likely to feel overwhelmed and unable to cope with the responsibility of raising a deaf or hard-of-hearing child. These children often fall victim to neglect and derogatory labelling from family members (Olusanya, 2008). As low socioeconomic status is attributed to residency in impoverished areas, deaf and hard-of-hearing children are at an increased risk for malnutrition, poor housing and sanitation, chronic illness, and lack of adequate healthcare and education (Swanepoel & De Beer, 2012). This subsequently hinders their development and growth.

1.4.1.4. The Macrosystem

The macrosystem consists of the micro-, meso-, and exosystems each of which is influenced by the morals, values, social class, ethnic group, and culture to which the developing child belongs (Bruyere & Garbarino, 2010; Visser & Moleko, 2012). Bronfenbrenner (1979) highlights the importance of understanding the social and psychological features of culture as it can influence processes which occur at the microsystem level (Algood et al., 2011). Cultural beliefs can influence parental perceptions about the meaning and causes deafness often being viewed as a disability (Kapitanoff, Lutzker, & Bigelow, 2000). Deafness in some cultures has been attributed to superstitious beliefs, 'evil forces', or atonement of sin (Olusanya, 2008). Parents and caregivers are then more likely to turn to traditional healers for a cure which often exposes their children to potentially harmful therapies (de Andrade & Ross, 2005; Olusanya, 2008). Deaf and hard-of-hearing children tend to be sent away to live with extended family members in rural villages due to the stigma attached to having a 'disabled' child (Olusanya, 2008).

By defining the various levels of the various levels of the environment, it is clear to see that these levels interact with and influence one another. Thus, it is important to understand the deaf and/or hard-of-hearing child at all levels of their environment in order to assist them in reaching their optimal development and to bring about systematic change. This proves important in early hearing detection and intervention (EHDI) programmes which aim to provide early hearing detection and subsequent therapeutic intervention services to the deaf and hard-of-hearing child.

1.5. Early Hearing Detection and Intervention (EHDI) and Universal Newborn Hearing Screening (UNHS) Programmes

The adverse consequences of late identified hearing loss on childhood development has been well-documented (Yoshinaga-Itano, Sedey, Coulter, & Mehl, 1998a; Yoshinaga-Itano, Apuzzo, 1998b; Moeller, 2000; Yoshinaga-Itano, 2004; Horn, Pisoni, & Miyamoto, 2006). In an attempt to combat the adverse developmental outcomes and to ensure timely transition from identification to intervention, the Early Hearing Dectection and Intervention framework was introduced (Moodley, 2016). The primary focus of any EHDI programme has been on the implementation of newborn screening (NHS) (Moodley, 2016). The earlies work on NHS dates back to when paediatric pioneer Dr Marion Downs implemented the first large scale NHS programme in 1963 (Hall, 2015). Her efforts in raising awareness of the implementation of EHDI on childhood development had led to the formulation of the first American diagnostic guidelines in 1982 (i.e., the Joint Committee on Infant Hearing (JCIH) Position Statement: Principles and Guidelines for Early Hearing Detection and Intervention) (Northern, 2015). Since then, many countries have adopted the recommendations laid out by the JCIH. The guidelines state that all infants should be screened within the first month of age, diagnosed by 3 months of age, and referred to an early intervention programme before the age of 6 months (JCIH, 2019). These guidelines have been adopted worldwide with NHS programmes being implemented in countries such as the United Kingdom (Kennedy, 2000); Australia (NSW Department of Health, 2011; Beswick, Driscoll, Kei, & Glennon, 2012); Spain (Benito-Orejas, Ramirez, Morais, Almaraz, & Fernández-Calvo, 2008), and Belgium (Vos, Senterre, Lagasse, & Levêque, 2015). The goal of these programmes is to identify infants with hearing loss in a quick and cost-effective manner, thus setting electrophysiological testing as a gold standard.

1.5.1. Electrophysiological Testing in UNHS

Audiological evaluations consist of behavioural, electroacoustic, and electrophysiological procedures (Jacob-Corteletti, Araújo, Duarte, Zucki, & Alvarenga, 2018). Technological advancements in neonatal hearing screening technology and the implementation of newborn hearing screening (NHS) programmes has provided more objectively-determined measures to identify infants who may be deaf or hard-of-hearing within the first few months of life (Cone-Wesson et al., 2000; Imam, El-Farrash, & Bishoy, 2013; Jacob-Corteletti et al., 2018). These electrophysiological measures consist of a set of examinations which record and analyse the physiological responses in the auditory system to determine hearing thresholds, locate lesions in the auditory pathways and diagnose retrocochlear injury (Bakhos, Marx, Villeneuve, Lescanne, Kim, & Robier, 2017). While several electrophysiological measures exist, automated otoacoustic emissions (OAE) and automated auditory response (AABR) technologies have been endorsed by the JCIH (2019) and the HPCSA (2018) for use in neonatal auditory screening.

1.5.1.1. Automated Otoacoustic Emissions (OAEs)

Otoacoustic emissions (OAEs) are low level sound waves emitted by the motion of the outer hair cells (OHC) of the cochlear in response to sound (Cunningham, 2011). OAE screening technology measures the functionality of the peripheral auditory system through a probe which is inserted into the external auditory canal (Bakhos et al., 2017). OAEs are subdivided into two types of measures: transient evoked otoacoustic emissions (TEOAE) and distortion product otoacoustic emissions (DPOAE). TEOAEs are produced by the OHC of the cochlear in response to brief clicks or other otoacoustic stimuli across a broad range of frequencies (McCreery, 2013). DPOAEs are induced by simultaneously presenting two tones

at different levels and frequencies (McCreery, 2013). The response, or DPs, are emitted by the cochlear at a third frequency which is then measured by the device and analysed (Nazir, Gupta, Mir, Jamwal, Kalsotra, & Singh, 2016). Different frequency combinations may elicit different responses from different regions in the cochlear (McCreery, 2013).

1.5.1.2. Automated Auditory Brainstem Response (AABR)

Auditory brainstem responses (ABRs) are electrophysiological responses to otoacoustic stimuli which is measured through electordes and transducers which are applied to the forehead (Neumann & Indermark, 2012). In comparison with OAE measures, AABR devices are able to measure the status of the peripheral auditory pathways which extend beyond the cochlear into the lower brainstem and thus allows for the detection of auditory neuropathy (Benito-Orejas et al., 2008).

These electrophysiological measures provide reliable and objective information regarding the underlying physiological activity in the auditory function (JCIH, 2019). he technology underlying the OAE and AABR devices make use of binomial automated response algorithms (i.e., “pass” or “refer”) and statistical software (e.g., the Quickscreen programme, Madsen-Gn, Otometrics, and DPGRAM software) to provide a statistical confirmation of the presence or absence of auditory activity (Benito-Orejas et al., 2008; MAICO Diagnostics, 2009; Nazir et al., 2016).

While both OAE and AABR measures have been successfully implemented in NHS programmes, differences between the two exist as the mechanisms they examine differ accordingly (Hall, 2015). A significant drawback of OAE testing is that it exclusively evaluates the functioning of the peripheral system. As a result, infants who may be at-risk for auditory neuropathy will be missed (Bezuidenhout, 2016). AABR testing accounts for this, as

it is able to measure both the functioning of the cochlear as well as the neurophysiological responses from the brain's response to sound (Hall, 2015). An additional disadvantage of OAE screening is that OAE signals tend to be highly sensitive to the presence of fluid and/or debris in the ear canal after birth (Kemp, 2002). It is therefore recommended that infants are only screened a minimum of 48 hours after birth (Benito-Orejas et al., 2008). However, due to shorter hospital stays particularly in public hospitals, babies who do not require neonatal care are being screened much earlier. This, in turn, increases the number of false-positives and a high number of referral rates (Benito-Orejas et al., 2008; Bezuidenhout, 2016). As AABR devices are sensitive to both cochlear and retrocochlear pathology, ABRs can be recorded after a few hours of birth when OAEs cannot (Kemp, 2002). As such, AABR screening is often used as a follow-up procedure if the infant failed the initial OAE test (Olusanya et al., 2005). However, AABR devices are costly to purchase and require a much longer examination time by an experienced audiologist., compared to OAEs which are relatively time efficient and simple to perform by both audiologists and non-audiological staff (Benito-Orejas et al., 2008). Both measures are highly sensitive to environmental factors such as ambient acoustic sound which often result in high failure rates (Hall, 2015; Van Dyk, Swanepoel, & Hall, 2015). It is thus important to understand the origins of each of the respective responses and limitations of each technique. By using the two as a two-step or "cross-check" procedure, it will significantly reduce the number of both false-positives and false-negatives (Hall, 2015). While this is ideal, it is difficult to implement in resource-poor or developing countries due to the high cost and maintenance of these devices. This has ultimately hindered UNHS in countries like South Africa (Bezuidenhout, 2016).

1.5.2. EHDI and UNHS in South Africa

The endorsement of the EHDI framework through the integration of the interdisciplinary systems of UNHS, diagnosis, and intervention has become the gold standard of care to children who are deaf and hard-of-hearing (Kamal, 2013). There has been much evidence with regards to the benefits of both UNHS and EHDI programmes on the achievement of developmental milestones, academic performance, and socioemotional and communicative functioning in developed countries (Khoza-Shangase, 2019, p. 73). Unfortunately, developing countries face numerous challenges in implementing both EHDI and UNHS programmes. In an attempt to spark the EHDI drive in South Africa, the Professional Board for Speech, Language, and Hearing Professions released a position statement providing specific guidelines with regards to EHDI in South Africa in 2007 (HPCSA, 2018). However, NHS is not mandated by the Department of Health, and there are additional challenges such as financial constraints, lack of manpower and resources, and poor implementation of NHS protocols by hospital staff (Bezuidenhout, Khoza-Shangase, De Maayer, & Strehlau, 2018). Due to these constraints, UNHS in South Africa, particularly in public hospitals, may be unfeasible to implement (Bezuidenhout, 2016; Kanji, 2016). As such, it has been recommended that a risk-based newborn screening approach should be used as an interim to UNHS. This risk-based approach should be offered to targeted groups who have one or more known risk-factors for hearing loss (Olusanya, 2011b; HPCSA, 2018).

Targeted newborn hearing screening (TNHS) aims to identify and screen any infants who are considered to be at-risk for having and/or developing a hearing loss based on established risk factors (Olusanya et al., 2005). Endorsing a targeted-based approach to newborn screening allows for the identification of deaf and hard-of-hearing children in areas whether UNHS programmes have not yet been implemented. Additionally, TNHS further allows for the identification and monitoring of children who may require additional medical

and/or audiological care (Kanji & Khoza-Shangase, 2019, p. 53) due to an underlying comorbid disorder (HPCSA, 2018, p. 18) and/or likelihood of developing a late-onset hearing loss. Despite the advantages of endorsing TNHS as a first step towards the implementation of a full UNHS programme, at least 50% of hearing loss occurs in babies who have no known cause or risk factors related to deafness (Durieux-Smith & Whittingham, 2000; Hyde, 2005). Consequently, TNHS results in a significant number of deaf and hard-of-hearing children being unidentified (HPCSA, 2018). An additional challenge of implementing TNHS are the risk factors with which the programme is based. Kanji and Khoza-Shangase (2019) argue that while the risk factors presented in the JCIH have been considered applicable for use internationally, context-specific risk factors have not yet been established. While it is estimated that half of congenital and early-onset hearing loss has been attributed to genetic causes, context-specific risk factors are important to establish as the risk factors may vary from country to country (Olusanya, 2011b). As such, risk factors applicable in developed countries may exclude potential risk factors related to those in developing countries such as infectious diseases (e.g., HIV/AIDS, TB, malaria) as well as additional risk factors related to socioeconomic status, culture, and poverty (e.g., maternal hypertension disorder, premature birth, and low-birth weight) (Khan et al., 2018). The HPCSA (2018) has provided a high-risk registry (HHR) which contains a recommended list of risk factors associated with neonatal hearing loss. These risk factors have been derived from the high-risk registry (HHR) proposed by the JCIH (2019) and modified to include the risk factors considered contextually relevant to South Africa (Kanji, 2016).

However, the context-specific risk factors are limited. This may be accounted for by the absence of an online data management system to log audiological data from screening initiatives in South Africa (Melton, 2008; Moodley & Storbeck, 2017). Data management involves the process of collecting, storing, and modelling medical data with the intent to

develop and implement future EHDI programmes (Moodley & Storbeck, 2017). A national data management system would not only provide a platform for infants and children to be monitored from diagnosis to intervention (Moodley & Storbeck, 2017), but for crucial information (such as risk factors) to be collated and modelled to determine the risk factors present specifically in a South African context. Due to a lack of funding, a shortage of trained/skilled staff, limited access to electricity, and poor internet connection in South Africa (Olusanya, 2008; Moodley & Storbeck, 2017) such a system would be difficult to implement at a national scale in both private and public healthcare.

Nevertheless, the identification and diagnosis of childhood hearing loss is a crucial task for preserving later development. It is therefore of benefit to the field of audiology to develop predictive tools to examine and identify this potential risk. Through the introduction of machine learning tools and techniques to audiology, large amounts of medical data can be explored and modelled and then used to predict diagnostic results on unseen medical data in order to support future decision-making (Alonso et al., 2018) especially in poor-resourced regions of South Africa.

1.6. Machine Learning in Healthcare

The term “machine learning” was coined in 1959 by American computer and artificial intelligence (AI) pioneer Arthur Samuel. With the ever-advancing nature of computational technologies, machine learning has increasingly become predominant in the field of research (Shew, New, Wichova, Koestler, & Staecker, 2019). The general goal of machine learning is to provide systems with the ability to automatically process large amounts of data in order to understand its basic structure. Moreover, it allows for this data to be fit into meaningful statistical models without being explicitly programmed (Serra, Galdi, & Tagliaferri, 2017;

Wallace, Noel-Storr, Marshall, Cohen, Smalheiser, & Thomas, 2017). Machine learning techniques are versatile in that they are able to derive models from large amounts of data without prior knowledge of the relationships between variables (van der Heide et al., 2019). Furthermore, these techniques are able to process large amounts of data with fewer statistical assumptions (van der Heide et al., 2019, p. 9409).

Borrowing from the disciplines of computer science, mathematics and statistics, machine learning was originally created as a mechanism to mimic human intelligent behaviour (Han, Pei, & Kamber, 2011). It has since become a valuable prediction tool in many fields with applications in childhood development and education (Kotsiantis, 2012; Ambili & Afsar, 2016; Amrit, Paauw, Aly, & Lavric, 2017), finance (Heaton, Polson & Witte, 2016), and in the allocation of scarce resources (Ware, 2018; Kube, Das, & Fowler, 2019). More commonly, machine learning has played a particularly important role in healthcare and medical research (Yoo et al., 2012; Luo, Wu, Gopukumar, & Zhao, 2016; Alonso et al., 2017).

The increasing popularity in the use of electronic health records to improve the effectiveness and efficiency of healthcare providers has resulted in petabytes of unstructured medical data (Jha et al., 2009). By analyzing this accumulated medical data through the use of machine learning techniques has the potential to increase research possibilities and enhance healthcare services (Luo et al., 2016). Machine learning may provide opportunities for the development of predictive models, reduction of risks, individualized treatment services, and discovering patterns of behaviour (Alonso et al., 2017). Successful applications of machine learning by medical researchers and organisations have helped to predict insurance fraud and abuse (Kose, Gokturk, & Kilic, 2015); epidemiology (Wiens & Shenoy, 2018); and in the prognosis, diagnosis, and treatment of diabetes (Kavakiotis et al., 2017),

cancer (Kourou et al., 2015), cardiovascular disease (Weng, Reps, Kai, Garibaldi, & Qureshi, 2017), and psychiatric disorders (Iliev, Dehghani, & Sagi, 2015; Alonso et al., 2018).

1.6.1. Machine Learning in Audiology

In the field of audiology, machine learning techniques have been used to extract and apply useful information to practical problems. A recent study conducted by Shew et al. (2019) used machine learning to develop a model to predict the presence of sensorineural hearing loss in inner ear pathologies using microRNA expression profiles. Machine learning models were developed to assist otologists in the diagnostic process of otoneurological diseases such as Meniere's Disease, benign positional and traumatic vertigo, vestibular neuritis, and sudden deafness (Juhola et al., 2001). Zhao et al. (2019) used support vector machines (SMV), multilayer perceptron (MLP), adaptive boosting (Adaboost), and random forest (RF) to demonstrate the feasibility of using machine learning to predict noise-induced hearing loss in industrial workers. In terms of amplification, machine learning as been applied to determine factors influencing which patients would most benefit from hearing aid fittings (Anwar, Oakes, Wermter, & Heinrich, 2010; Panchev, Anwar, & Oakes, 2013). Although machine learning techniques have been applied to various subdomains in audiology, the use of these techniques specifically in the subdomain of neonatal and early childhood deafness has been underutilized.

Despite the advantages of machine learning, studies in the area of neonatal and early childhood deafness have mostly been conducted using traditional quantitative and qualitative techniques to analyse the risk factors associated with deafness (Yoshikawa, Ikeda, Kudo, & Kobayashi, 2004; Martinez-Cruz, Poblano, & Fernández-Carrocera, 2008; Kanji & Khoza-Shangase, 2012; Beswick et al., 2013; Swanepoel, Johl, & Pienaar, 2013; Kuschke,

Goncalves, & Peer, 2018; Kanji & Khoza-Shangase, 2019), aetiology of deafness (Admiraal & Huygen, 1999; Derekoy, 2000; Egeli et al., 2003; Riga et al., 2005; Gruss, Berlin, Greenstein, Yagil, & Beiser, 2007; Kim, Choi, Han, & Choi, 2016), views of healthcare workers and parents of deaf and hard-of-hearing children (Moodley, 2012; Moodley & Storbeck, 2017; Davids & de Jager, 2018; Khan et al., 2018; Bhamjee et al., 2019; Gardiner, Laing, Mall, & Wonkam, 2019), early hearing detection and intervention (EHDI) services (Swanepoel, Storbeck, & Friedland, 2009; Findlen, Hounam, Alexy, & Adunka, 2019; Khoza-Shangase, 2019), and newborn hearing screening (NHS) programmes (Ciorba et al., 2008; Colgan et al., 2012; Olusanya et al., 2005; Gabriel et al., 2020; Bezuidenhout, 2016).

1.6.2. Extending Qualitative Research in Neonatal and Early Childhood Audiology with Machine Learning

An explicit comparison between machine learning and quantitative methods are difficult to achieve as although both methods seek to create models from data, each does so for a different purpose (Fawcett & Hardin, 2017). A common purpose in quantitative research is to draw formal conclusions about the relationship between variables to contribute scientifically to the field of study (Fawcett & Hardin, 2017). The aim is to provide a better understanding about the phenomenon and provide a framework with which better decisions and planning can be made (Fawcett & Hardin, 2017; Stewart, 2019). Quantitative research is rarely concerned with the prediction of future data, a common goal in machine learning (Stewart, 2019). Even if quantitative models are used to determine whether a causal relationship between variables exist, the relationship is evaluated by significance and the interpretation is final. In contrast, the main purpose of machine learning is to create a model which can generate the best repeatable prediction so that decisions can be made automatically and applied practically (Stewart, 2019).

Several studies in audiology have examined the risk factors which are associated with childhood hearing loss in order to understand the implications that it has for the planning and implementation of early identification and intervention services (Kanji, 2016). The current study aimed to extend on previous work by identifying and examining the risk factors present in a cohort of South African children and applying it to two classification models to predict the likelihood of high-risk children having or developing a hearing loss. Furthermore, the study sought to contribute to the development of a high-risk registry (HHR) for the South African population by examining the risk factors which have the most predictive influence on hearing loss.

1.7. The Current Study

1.7.1. Study Aims

The main aim of the current study was to use predictive modelling to predict the occurrence of hearing loss in high-risk neonates. This aim is two-fold:

1. To develop a model which predicts the occurrence of hearing loss in high-risk neonates;
2. To demonstrate the feasibility of using machine learning in predicting the future occurrence of hearing loss in high-risk neonates.

1.7.2. Study Objectives

The above-mentioned aim was achieved through the following objectives:

- Build and train two predictive models using the Random Forest (RF) classifier and Logistic Regression to predict the occurrence of hearing loss in high-risk infants.
- Test the trained models to predict the occurrence of hearing loss in neonates who have not yet been diagnosed with a hearing loss but may or may not present with one or more associated risk-factors.

1.7.3. Research Question

Based on the increase in popularity of the use of machine learning techniques as clinical risk prediction models in the medical, and more specifically, audiological domains, the following research question has been answered: Can the occurrence of neonatal hearing loss be predicted in high-risk neonates using machine learning techniques?

Chapter 2: Methodology

Examining the contribution of the variables related to neonatal hearing loss requires a dynamic approach which is firmly rooted in both quantitative and machine learning epistemologies. The inclusion of both quantitative and machine learning epistemologies increases the possibility of objectively identifying infants who are at risk for having a hearing loss (whether it is congenital or acquired), while reducing the subjective bias of healthcare workers where the risk of hearing loss going undetected may be high. This chapter provides a description of the adopted research design and analysis, and data and patient record characteristics. A thorough description of the data processing and statistical and machine learning analyses procedures are presented. The chapter ends with a discussion of ethical considerations adhered to for the duration of the study.

2.1. Dataset Description

2.1.1. *Sampling Methods*

A secondary, convenience sampling method was utilized. Data from two pre-existing secondary datasets were used to create predictive models to predict hearing loss in high-risk neonates. Convenience sampling is a non-probability sampling method where members of the target population are selected for inclusion in the sample study due to close proximity, availability, and accessibility (Etikan, Musa, & Alkassim, 2016). It is noted that there are several relating to the secondary use of data. With the data not have been collected for the specific purpose of the current study, the methodology of the initial data collection, data accuracy, period of data collection, purpose for which the data was collected, and the content of data needs to be evaluated for reliability and validity (Tripathy, 2013). This was addressed by thoroughly examining the provenance of the secondary datasets through careful inspection

of each programme's strategic research framework which included, but was not limited to: research aims objectives, conceptual framework, and research strategy and design. On the grounds of the manner in which the data was collected, the nature of the variables that were to be used in the current study, and the prominence of both organisations, there were no cause for concern regarding the reliability and validity of the data used in the current study.

2.1.2. Secondary Datasets: HI HOPES and Netcare Hospital Group

The secondary datasets consisted of: (1) longitudinal data from an early hearing intervention programme and; (2) data collected through a newly implemented universal newborn screening programme. It is to be noted that both of these programmes are the first of their kind to be launched in South Africa (Storbeck & Young, 2016; Netcare, 2019).

2.1.2.1. HI HOPES Dataset.

Launched in Gauteng 2006, HI HOPES (Home Intervention Hearing and language Opportunities Parent Education Services) is the first home-based family-centred early intervention programme in South Africa (Storbeck & Young, 2016). Based on the SKI-HI Model of Early Intervention (SKI-HI Institute, 2004), the HI HOPES programme provides support to families of deaf and hard-of-hearing children (aged 0 to 3) with unilateral and bilateral hearing loss ranging from mild to profound (Storbeck & Young, 2016). The HI HOPES programme aims to empower families by providing them with a holistic understanding and skillset to advocate and make informed decisions for their child whenever deemed necessary (Storbeck & Young, 2016). The programme is non-discriminative, providing services to all families regardless of healthcare status (public or private) at no additional cost (Storbeck & Young, 2016).

The dataset consists of structured longitudinal data pertaining to deaf and hard-of-hearing children and their families who were enrolled in the HI HOPES programme over a 12-year period (September 2006 – December 2018). The dataset is comprised of a comprehensive history of each individual child and includes family demographic and socioeconomic variables. All information pertaining to each child and their family is updated at regular intervals throughout the child's enrollment in the programme (Storbeck & Young, 2016). While the dataset consists of numerous variables related to each child and their families, the current study will only make use of the associated risk factors for hearing loss listed by the HPCSA (2018) and birthing information, namely: birth type (natural, elective caesarian, emergency caesarian) and gestational age (to determine prematurity).

Ethical approval for the creation and analysis of the HI HOPES project and relevant data has previously been obtained by the key gatekeepers at the Centre for Deaf Studies, University of the Witwatersrand, from the Wits Human Research Ethics Committee (Non-Medical) for creation of the dataset (Protocol number 2007 ECE20).

2.1.2.2. Netcare Hospital Group Dataset.

Netcare Hospital Group, one of the leading private healthcare groups in South Africa, launched a UNHS programme in 37 of its hospitals spanning across 7 provinces on the 1st of June 2019 (Netcare, 2019). Newborn hearing screening is offered to every child born at participating Netcare hospitals before being discharged in order to follow the 'universal' hearing screening principle. The UNHS programme is linked to a maternity passport which offers basic OAE screening. Babies referred to the neonatal intensive care unit (NICU) are offered both OAE and AABR screening procedures. All screening data is logged onto a data management application and includes birthing information, parent and child details, screening methods and results, and risk factors associated with hearing loss. The current study will only

make use of the associated risk factors for hearing loss listed by the HPCSA (2018) and relevant birthing information such as birth type (e.g., natural, elective caesarian, emergency caesarian) and gestational age. It is important to note that regardless of whether the infants included in the dataset passed or failed their initial screening test, none have any known diagnosis of a hearing loss at this stage.

Permission to use the data for secondary analysis of this dataset was obtained from the key gatekeepers at the Centre for Deaf Studies, University of the Witwatersrand. Ethical approval for the collection of this data was previously obtained from the Wits Human Research Ethics Committee (Medical).

2.1.3. Patient Records

The total sample included in the current study comprised 12 044 patient records of neonates who were part of the HI HOPES programme (89%, n = 10 700) or the Netcare UNHS programme (11%, n = 1 344). All children who formed part of the HI HOPES programme or the Netcare UNHS programme and who had complete data for all predictor and outcome variables were eligible to be included in the current study.

2.2. Research Design

For the purpose of the current study, a nonexperiemental, predictive modeling design was employed. Predictive modeling, as defined by Geisser (1993) refers to the process of developing a mathematical model to best predict the probability of an even or outcome occurring. The practice of predictive modelling is concerned with developing a model in such a way that future events or behaviour can be accurately forecasted. It involves the assessment of variables at one point in time so as to predict the occurrence of a phenomenon associated with those variables at a later point in time (Kuhn & Johnson, 2013). As the current study was concerned with creating a model which predicts the likelihood of a hearing lows occurring in high-risk neonates, the chosen design, which uses predictive modelling, aligns with the research aims and objectives.

2.2.1. Outcome and Predictor Variables

Independent variables are variables that cause, affect, or influence changes in a phenomenon whereas the dependent variable is the outcome or results of the influence of the independent variable (Creswell, 2003). Nonetheless, the term predictor is often used in nonexperimental research to refer to a variable that can predict another variable (i.e., the independent variable) (Flannelly, Flannelly, & Jankowski, 2014). The term predictor is useful as it does not imply that the predictor causes change in the outcome variable even though it may (Flannelly et al., 2014). For the purpose of the current study, the independent variables will be referred to as the predictor variables and the dependent variables will be referred to as the outcome variables.

The dichotomous outcome variable was hearing status (i.e., hearing or deaf). The predictor variables used to predict hearing status included: birth type (natural, elective

caesarean, emergency caesarean), gestational age, premature birth, family history, extracorporeal membrane oxygenation (ECMO), hyperbilirubinemia requiring exchange transfusion, in-utero infections, craniofacial anomalies, physical findings (such as white forelock and distinctive facial features), syndromes associated with hearing loss or progressive/late-onset hearing loss, cultural-positive postnatal infections, maternal and/or infant HIV infection, meningitis, and exposure to ototoxic medication. All predictor variables, save for gestational age, were categorical (e.g., "Infant has a family history of hearing loss" – Response: Yes or No).

2.2.2. Classification: A Machine Learning Technique

Classification is a popular supervised learning technique which has been extensively used in the medical field to aid in the diagnosis, prognosis, and treatment plan of health outcomes (Harper, 2005; Yoo et al., 2012). Classification involves the process of deriving models, called classifiers, which predict class labels of given data points (Asiri, 2018). These 'classes' are the attributes in a dataset which are of most interest and is more commonly known as the dependent variable in statistics (Yoo et al., 2012). Classification occurs a two-step process in which: (1) a model is constructed by analyzing training data containing class labels and classification rules (the learning step) and; (2) the model is examined for accuracy or its ability to classify unseen data (the classification and/or training step) (Han et al., 2011; Yoo et al., 2012). While predictive accuracy it an important criterion for a good classification model, the model's ability to provide an understanding of the underlying predictive structure of the data is of equal importance. In neonatal audiology, finding out which key characteristics and risk factors contribute to the risk of infants developing a hearing loss would not only provide assistance in classifying infants some certainty into risk groups, but

to more generally advance the knowledge and understanding of the causes and/or risks of hearing loss.

Many different classification techniques are in existence such as Artificial Neural Networks (ANNs), Naïve Beyes classifier, Decision Trees (DT), Support Vector Machine (SVM), Association Rules (ARs), Regression Models, and Discriminant Analysis (DA). However, their merits and practicality for usage especially in the medical field remain unclear (Harper, 2005). Two classification techniques namely, Random Forest and Logistic Regression, have been considered in the current study to evaluate their relative performance in predicting hearing loss in high-risk neonates. A brief synopsis of each is provided below.

2.2.2.1. Logistic Regression.

In general, regression analysis seeks to predict the relationship between a response variable and one or more predictor variables in the presence of random error (Abdulqader, 2017). In situations where the response variable is dichotomous (binary), logistic regression is used. Logistic regression uses the theory of binomial probability to model the chance of an outcome occurring based on individual variable characteristics (Sperandei, 2014). It applies the maximum-likelihood estimation (MLE) algorithm to search for the best coefficients which would result in a model which would predict a value very close to 1 (i.e., the event/outcome occurring) and 0 (i.e., the event/outcome not occurring). By seeking the best coefficient estimates of the logistic regression, the error predicted by the model is minimized.

Logistic regression has been successfully used in many healthcare instances (Abdulqader, 2017) and is an especially powerful tool in epidemiologic studies as it allows for multiple predictor variables to be analysed simultaneously while reducing the effect of confounding factors (Sperandei, 2014).

2.2.2.2. Random Forest.

Random Forest (RF) is a classification technique which uses the ensemble machine learning algorithm to aggregate the predictive ability of multiple classifiers (Husain & Xin, 2016). Put simply, it builds a "forest" consisting of multiple random decision tress where each tree in the forest produces a class prediction after searching for the best features amongst a random subset of features present (Donges, 2020). The class prediction with the most votes in the entire forest becomes the overall models class prediction (Yiu, 2019). This ensemble method results in a more comprehensive classifier (Hussain & Xin, 2016).

An important feature of the Random Forest classifier is its ability to calculate variable importance. The classifier can analyse each attribute and reveal the importance of the attribute in making the correct classification of the model (Livingson, 2005). This allows researchers to filter out unnecessary attributes which could save time during data collection and data analysis (Livingson, 2005, p. 2).

The Random Forest approach has shown good accuracy in overall performance (Sarica, Cerasa & Quattrone, 2017). In general, it produces a higher prediction accuracy, is robust to overfitting, and is considered more stable in the presence of outliers and high dimensional parameter spaces than any other classification technique (Hussain & Xin, 2019).

2.3. Procedure

Prior to conducting statistical and predictive analyses, pre-processing techniques were conducted on each individual dataset in Microsoft Excel to ensure accuracy, completeness, and consistency in the data. Any missing values, noise, and inconsistencies which may bias the final dataset and consequently result in statistical and classification underperformance (Husain et al., 2016) were dealt with as described below. Once both datasets were verified

and validated, the datasets were carefully integrated so as to reduce and avoid redundancies and inconsistencies in the resulting dataset. The data was then exported into R Version 1.3 (RStudio.com, 2020) for analysis.

2.3.1. Handling Missing Values

Missing values are a universal problem in the analysis of large datasets across many research domains (Abidin, Ismail, & Emran, 2018; Schmidt, Niemann, & Trzebiatowski, 2015). Missing values are very common and occur due to data corruption, measurement error, non-response from participants or failure to record data (Kumar, 2020). Unfortunately, in medical data, missing values are unavoidable. This poses a problem in the results of statistical and machine learning methods which often assume completeness of data (Kumar, 2020; Ding & Simonoff, 2006). Missingness in data imposes undesirable effects on statistical and machine learning results by introducing the element of uncertainty by distorting the final results and causing a biased statistical analysis (Schmidt et al., 2015; Abidin et al., 2018).

Previous researchers have proposed several ways in which to manage missing data (Liu, White, Thompson, & Bramer, 1997; Batista & Monard, 2003; Kim & Yates, 2003; Zhang, Qin, Ling, & Sheng, 2005; Ding & Simonoff, 2006; Saar-Tsechansky & Provost, 2007; Schmidt et al., 2015; Abidin et al., 2018). The most traditional method of dealing with missing values is to discard the missing instances (Ding & Simonoff, 2006; Saar-Tsechansky & Provost, 2007; Schmidt et al., 2015; Abidin et al., 2018). Through this method, cases with missing values are omitted leaving the remaining data complete and statistical and machine learning techniques can be applied without further issues (Schmidt et al., 2015; Abidin et al., 2018).

The challenge of missing or unknown data affecting the accuracy of the models in the current study was addressed by excluding cases where large volumes of data was missing on key predictors. As one of the objectives of the current study is to assess the performance of using machine learning methods in predicting the likelihood of a hearing loss occurring in high-risk neonates, this strategy was deemed appropriate (Saar-Tsechansky & Provost, 2007). While handling missing data with this method is the easiest and thus default option in most statistical and machine learning methods, it may lead to a loss of critically important information and may bias parameters and estimates (Abidin et al., 2018).

2.3.2. Handling Class Imbalance

After analysing the cleaned dataset, class imbalance was detected. Class imbalance occurs when the number of instances in one class, usually the one which is of more interest, or the 'positive class' is significantly outnumbered by the majority class (Zhu et al., 2019). In the current study, the deaf and hard-of-hearing class was greatly outnumbered by the hearing class by 1:8. Class imbalance poses a major challenge for classification techniques as it can decrease the effectiveness of the classification model (Chen et al., 2004). As such, classification models ran on imbalanced data often results in a high overall accuracy rate but a low accuracy rate for the positive or minority class (Chen et al., 2004; Zhu et al., 2019). This is because classification models are more biased towards the majority class as they aim to minimize the overall error rate rather than considering the importance of the minority class (Zhu et al., 2019). Hence, a process for handling imbalanced data before constructing and running the two classification models proposed in the current study was required. Common solutions to handle class imbalance include random oversampling which, in effect, resamples the data present in the underrepresented class and, random under-sampling which downsizes

the over-represented class by removing training examples from that class (Kotsiantis et al., 2006).

Based on the ratio of the positive to negative instances in the data used for the current study, the under-sampling method was eliminated as the positive class (i.e., deaf and/or hard-of-hearing) was substantially lower than the negative class (i.e., hearing). This would have resulted in a drastic decrease in sample size and a loss of potentially important information. Therefore, *ROSE* (version 0.0-3), a popular package for binary imbalanced learning, was used to resample the data of the underrepresented class (Lunardon et al., 2014). This process was executed by the *ROSE* package (version 0.0-3) by randomly oversampling the data with replacement until the specified sample size reached (Lunardon et al., 2014).

2.3.3. Statistical and Machine Learning Analyses

Descriptive data analyses were conducted to describe the neonate profile (i.e., hearing status, gender, programme, and healthcare status), mode of delivery, gestational age, and risk factors associated with hearing loss. The categorical variables were presented as frequencies and percentages while the continuous variable (i.e., gestational age) was presented by the mean, standard deviation, and range.

The association between the HPCSA risk factors with the addition of mode of delivery, prematurity, and gestational age and hearing loss was determined using the Chi-Squared Test of Contingencies (X^2). A simple logistic regression model was computed to determine the relationship between gestational age and hearing loss. Predictor variables with a p-value of < 0.05 were included in final classification models.

Two classification models namely, logistic regression and random forest, were built with the dataset consisting of both hearing and deaf and/or hard-of-hearing children who may

or may not have presented with the risk factors associated with neonatal hearing loss. Each child was given a class label which indicated their hearing status, thus resulting in two distinct classes namely, “HL = 0” for hearing children and “HL = 1” for deaf and/or hard-of-hearing children. To construct, test, and compare the two classification techniques, the balanced study dataset was split into a ‘training’ set in which the models were derived, and a ‘testing’ set in which the models were applied to and tested. The sample was randomly split into two parts for training (70%) and testing (30%). The development of the classification models in the training set, and the application of the classification models to the testing set, was conducted in R using the library packages *caret* (version 6.0-86) and *randomForest* (version 4.6-14). Model performance for both models were assessed using accuracy, specificity, and sensitivity scores. Internal validation of the models was assessed using 10-fold cross-validation with 10 repeats (Moons et al., 2015). Cross-validation was used to adjust for overfitting in the predictive ability of each model (Moons et al., 2015).

2.4. Ethical Considerations

Permission to conduct the current study using the datasets described above was obtained from the head researchers at the Centre for Deaf Studies and Netcare Hospital Group respectively. Ethical clearance for the current study was obtained from the Wits Human Research Committee (Medical) (protocol number: M200516 MED 20-05-051) (see Appendix B) upon acceptance of the current study by the Faculty of Humanities.

Both datasets contain sensitive information pertaining to a vulnerable population as the patient records included neonates who are underage, with some being deaf and/or hard-of-hearing. However, previous ethical clearance has previously been obtained for the creation and analysis of both datasets respectively as described above. Moreover, in conforming with

the ethical principle of anonymity and confidentiality, the data provided for secondary data analysis has been pseudo-anonymized with numbers being used as infant identifiers. Only the key gatekeepers have access to all infant information. Additionally, any data in the principal researcher's possession will be retained only as long as it is necessary to complete the current research project and possible publications which may arise from it.

Chapter 3: Results

This chapter presents the data analysis and results from the current study followed by a discussion of the findings presented in the next chapter. The data analysed to develop predictive models which predict the occurrence of childhood hearing loss based on the risk factors associated with childhood hearing loss.

3.1. Descriptive Analyses

3.1.1. Patient Records

A combined total of 25 992 patient records were initially included in the current study. Of these 25 992 records, 13 948 were excluded from the total sample due to high volumes of missing or unknown data. This left 12 044 patient records eligible for inclusion in the final sample. Of the 12 044 patient records, 10 700 (88%) neonates were hearing and were part of the Netcare UNHS programme, while 1 344 (11%) neonates were deaf and/or hard of hearing and were part of the HI HOPES programme. For the purpose of the current study, the patient records from both datasets were combined into a single sample and thus the demographics described further form part of the combined sample. The sample consisted of more males with 6 204 (51.5%) of the neonates being male and 5 840 (48.5%) being female. Ten thousand eight hundred and fifty-three (90%) of the neonates had access to private healthcare whereas 1 164 (9.7%) of the neonates having access to public healthcare. The remaining twenty-seven neonates (< 1%) had access to both private and public healthcare.

3.1.1.1. Mode of Delivery

Three thousand and ninety-three neonates (26.5%) were delivered via natural birth and 8 851 (73.5%) were delivered via caesarean section. Of the 8 851 delivered via caesarean section, 6 408 (72%) were delivered via elective caesarean, 2 385 (27%) were delivered via

emergency caesarean, and fifty-eight participants (1%) were delivered via unspecified caesarean section.

3.1.1.2. Gestational Age

The mean gestational age was 38.04 weeks ($SD_{GestationalAge}$ = 2.22, $Range_{GestationalAge}$ = 20 – 48 weeks). To analyse the relationship between hearing loss and prematurity, the neonates were classified into preterm (i.e., born prior to the completed 37th week of gestation), full-term (i.e., born between the 37th and 41st week of gestation), and post-term (i.e., born after the 41st week of gestation) groups (Fleischman, Oinuma, & Clark, 2010; Jacob et al., 2017). One thousand six hundred and nineteen neonates (13.4%) were classified as preterm, 10 335 (86%) were classified as full-term, and 92 were classified as post-term (< 1%).

3.1.2. Risk Factors

The current study made use of the risk factors defined by the HPCSA (2018) with the addition of mode of delivery (i.e., natural, elective caesarean, emergency caesarean), prematurity, and gestational age in all subsequent statistical and machine learning analyses. The frequencies and percentages of the HPCSA risk factors present in the cohort are presented in Table 1. The most frequently reported risk factor in the cohort was ototoxic medication with (24.06%) of children having this risk factor. Family history (19.45%) and ECMO (18.07%) were the next most frequently reported risk factors in the cohort. The next most common risk factors were cultural-positive postnatal infections (9.91%), meningitis (8.64%), maternal and/or infant HIV infection (5.03%), physical findings (3.92%), craniofacial anomalies (3.29%), in-utero infection (3.23%), syndromes associated with childhood hearing loss (2.38%), hyperbilirubinemia requiring exchange transfusion (1.75%). The least reported risk factor was neurodegenerative disorders (0.26%) with only 5 children

having this risk factor. Due to the low presence of neurodegenerative disorders in the study, this risk factor was dropped from any further analyses.

Table 1

Frequencies and Percentages of HPCSA Defined Risk Factors Present in Sample (n = 1887)

HPCSA Defined Risk Factors Present in Sample	N	%
Family History of Permanent Childhood Hearing Loss	367	19.45
Extracorporeal Membrane Oxygenation (ECMO)	341	18.07
Hyperbilirubinemia Requiring Exchange Transfusion	33	1.75
In-utero Infections	61	3.23
Craniofacial Anomalies	62	3.29
Physical Findings	74	3.92
Syndromes Associated with Hearing Loss or Progressive/Late-Onset Hearing Loss	45	2.38
Neurodegenerative Disorders	5	0.26
Cultural-Positive Postnatal Infections	187	9.91
Meningitis	163	8.64
Maternal and/or Infant HIV Infection	95	5.03
Ototoxic Medication	454	24.06

3.1.3. Statistical Analyses

Several Pearson's Chi-Square (χ^2) Test of Contingencies (with $\alpha = .05$) were computed to evaluate the association between each of the individual risk factors and the occurrence of hearing loss. The results from the χ^2 analysis presented in Table 2 showed

significant correlations between each birth type (i.e., natural, elective caesarean, emergency caesarean), prematurity, family history, ECMO, in-utero infection, craniofacial anomalies, physical findings, syndromes associated with hearing loss, cultural-positive postnatal infections, meningitis, maternal and/or HIV infection, and ototoxic medication. However, the association between hyperbilirubinemia requiring exchange transfusion ($\chi^2(1) = .53$, $p = .46$) was nonsignificant. As a result, hyperbilirubinemia was excluded from any further analyses.

Table 2

Pearson's Chi-Square (χ^2) Test of Contingencies Associations Between Risk Factors and Hearing Loss

	df	χ^2	*p*
Natural Birth	1	1459.53	<.001
Emergency Caesarian	1	16.62	< .001
Elective Caesarian	1	1124.09	< .001
Premature Birth	1	559.63	< .001
Family History	1	981.24	< .001
ECMO	1	20.66	< .001
In-Utero Infection	1	255.30	< .001
Craniofacial Anomalies	1	212.90	< .001
Physical Findings	1	441.58	< .001
Cultural-Positive Infections	1	1060.64	< .001
Syndrome	1	27.11	< .001
Meningitis	1	1140.09	< .001
Maternal and/or Infant HIV Infection	1	13.9	< .001
Ototoxic Medication	1	12.92	< .001

Hyperbilirubinemia Requiring Exchange Transfusion	1	.53	.46

To determine the relationship between hearing loss and gestational age, a simple logistic regression analysis was conducted. The omnibus model for the logistic regression analysis was statistically significant χ^2 (1) 131.02, p = .000. Coefficients are presented in Table 3.

Table 3

Logistic Regression Coefficients for Gestational Age as a Predictor of Hearing Loss

	b	*SE (b)*	*p*	*Exp (B) [95% CI]*
Constant	2.83			
naturalYes	-0.13	0.01	.000***	0.88 [0.86, 0.90]

3.1.4. Classification Models: Logistic Regression and Random Forest

3.1.4.1. Logistic Regression

A logistic regression classification model was constructed to predict the probability of hearing loss based on risk factors associated with childhood hearing loss. Assumption testing prior to constructing the model did not indicate any violations. The model was constructed in a two-step process where: (1) the model was built using the training set and then, (2) applied to the test data to examine the model's accuracy and performance.

The omnibus model for the logistic regression analysis was statistically significant χ^2 (15) = 8935.14, $p < .000$, Cox and Snell's $R^2 = .45$, Nagelkerke's $R^2 = .60$. The classification model had an overall accuracy of 92% in predicting hearing loss in children. The proportional reduction in error (PRE) rate was .90, indicating a 90% reduction in error when predicting hearing loss. Sensitivity and specificity rates indicated that the model was 74% accurate in predicting children who are more likely to have a hearing loss and 84% accurate in predicting children who are less likely to have a hearing loss, respectively.

Internal validation of the model using a 10-fold repeated cross-validation with 10 repeats found the model's accuracy rate was 92%. The cross-validated model remained a good fit for the data with the validated model ($p < .000$) and model coefficients remaining statistically significant ($p < .05$). Coefficients are presented in Table 4.

As demonstrated in Table 4, family history, ECMO, infection, craniofacial anomalies, physical findings, syndrome, cultural-positive postnatal infections, meningitis, and ototoxic medication were the only predictors which significantly improved the model's predictive capability. infection (OR: 9.93; CI [4.34, 2.40]) Premature birth (OR: 7.36; CI [5.38, 1.01]) ECMO (OR: 9.60; CI [3.73, 2.09]), , craniofacial anomalies (OR: 6.54; CI [3.89, 1.67]), and meningitis (OR: 5.41; CI [1.33, 1.42]) were the top five variables which significantly improved the model's prediction with odds ratios over 5. That is, children who had either prematurity, ECMO, infection, craniofacial anomalies, and meningitis were over 5 times more likely to develop a postnatal hearing loss than children without any of these risk factors.

Table 4

Predictor Coefficients for the Logistic Regression Model Predicting Hearing Loss

	b	*SE (b)*	*p*	*Exp (B)* *[95% CI]*
Intercept	13.13	190.06	.94	5.03 [1.68, 1.52]
naturalYes	-15.73	190.06	.93	1.48 [7.02, 7.18]
electiveCYes	-18.49	190.06	.92	9.32 [4.31, 4.51]
emergencyCYes	-17.56	190.06	.93	2.35 [1.09, 1.14]
prematureYes	2.00	0.16	.000**	7.36 [5.38, 1.01]
gestationalAge	0.03	0.02	.25	1.03 [9.80, 1.08]
familyHistoryYes	2.76	0.178	.000***	1.58 [1.12, 2.24]
ECMOYes	-2.34	0.44	.000***	9.60 [3.73, 2.09]
infectionYes	2.30	0.43	.000***	9.93 [4.34, 2.40]
craniofacialYes	1.88	0.47	.000***	6.54 [3.89, 1.67]
physicalYes	2.62	0.53	.000***	1.37 [5.12, 2.85]
syndromeYes	1.04	0.57	.05*	2.83 [8.79, 8.36]
culturalYes	1.03	0.63	.05*	2.80 [1.39, 9.37]
meningitisYes	4.00	0.73	.000***	5.41 [1.33, 1.42]
HIVYes	-0.06	0.44	.89	9.41 [3.80, 2.14]
ototoxicYes	-0.84	0.29	.001**	4.33 [2.34, 7.49]

Note: CI = Confidence Interval

** p significant at .05*

*** p significant at .001*

*** p significant at .000*

3.1.4.2. **Random Forest**

A logistic regression classification model was constructed to predict the probability of hearing loss based on risk factors associated with childhood hearing loss. The model was constructed in a two-step process where: (1) the model was built using the training set and then, (2) applied to the test data to examine the model's accuracy and performance.

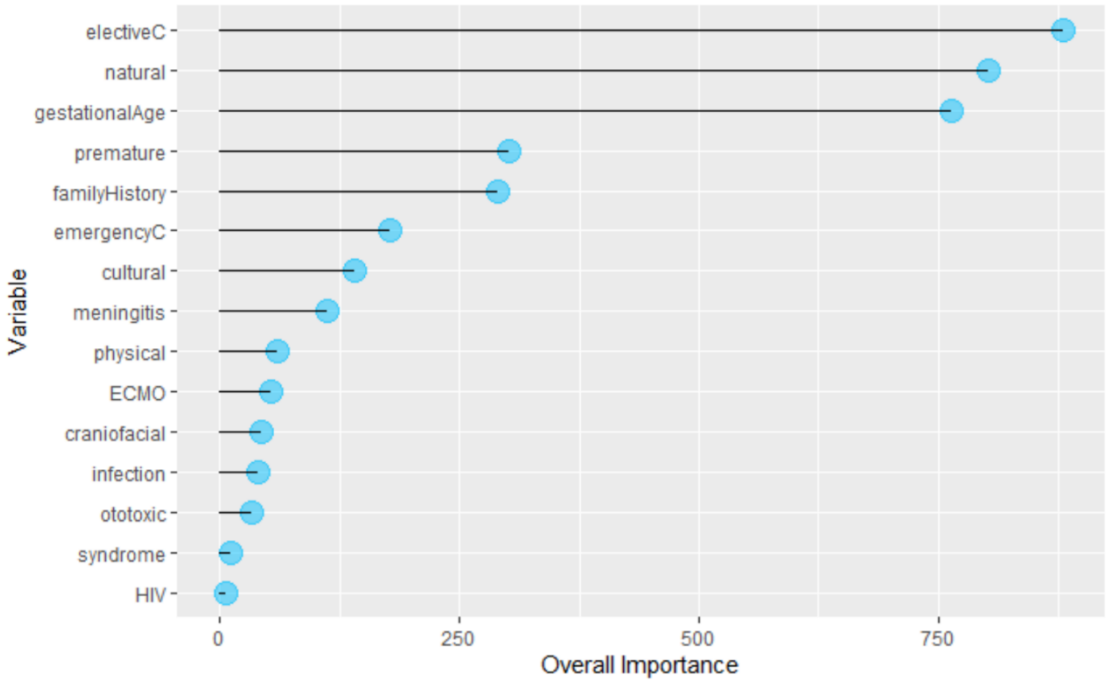
The random forest classification model had an overall accuracy rate of 88% in predicting hearing loss in children with sensitivity and specificity rates indicated that the model was 72% accurate in predicting children who are more likely to have a hearing loss and 90% accurate in predicting children who are less likely to have a hearing loss, respectively. The out-of-bag error rate (OOB) was 15.2%.

Internal validation of the model using a 10-fold repeated cross-validation with 10 repeats indicated that the model's best predictive accuracy rate would be obtained by tuning the number of predictors chosen at each split (*m*) to *mtry* = 8. Recalibration of the model's parameters to *mtry* = 8 and *ntree* = 100, found the model's overall accuracy rate remained the same at 88%. However, the out-of-bag error rate decreased slightly from 15.2% to 14.8%. Nonetheless, the cross-validated model remained a good fit for the data ($p < .000$).

To inspect which variables in the model had the most predictive power, variable importance statistics were run. As illustrated in Figure 2, all three modes of delivery (i.e., natural, elective caesarean, and emergency caesarean), gestational age, premature birth, family history, cultural-positive infections, and meningitis were important predictors in predicting hearing loss in children.

Figure 2

Variable Importance Plot of Predictors in Random Forest Risk Model



Chapter 4: Discussion

This chapter provides a discussion of the main findings from the current study in accordance with the specific aims and objectives of the study with reference to relevant literature in order to substantiate the findings from the current study. Limitations in the research methodology and design is also discussed, and recommendations are made for future research.

Early identification and intervention of childhood hearing loss is crucial to the well-being of both children and their families. Early hearing detection and intervention (EHDI) services have become an important goal in audiology to avoid or significantly minimise the effects of congenital and/or early-onset/progressive childhood hearing loss (Swanepoel, 2009; Khoza-Shangase, 2019). Several studies have shown the far-reaching effects of the late identification of childhood hearing loss on the achievement of milestones in the interrelated domains of language, cognition, and socio-emotional development (Moeller, 2000; Olusanya, 2008; Swanepoel, 2009; Kanji, Khoza-Shangase, Ballot, 2010; Khoza-Shangase, 2019). These effects may lead to ramifications in future scholastic and vocational achievement (Olusanya, 2008; Swanepoel, 2009; Kanji et al., 2010).

Evidence from the developed world has documented the positive effects of EHDI programmes not only on the domains of development but with regards to the interaction between parent and child resulting in more secure attachment bonds and a supportive family environment (Khoza-Shangase, 2019). This, in turn, results in an overall improvement in quality of life (Khoza-Shangase, 2019). However, the implementation of EHDI, particularly the endorsement of UNHS, faces many challenges especially within developed countries such as South Africa where manpower, financial aid, and technological resources are lacking. Consequently, targeted newborn hearing screening (TNHS) has been recommended as an interim approach in context such as South Africa where UNHS cannot yet be feasibly implemented. Despite the defined a list of risk factors found to be associated with hearing

loss proposed by the JCIH (2007), the risk factors applicable to developing countries such as South Africa are absent. While the HPCSA (2018) has adapted this list based on contextually identified findings in the South African population, these studies have evaluated hearing loss risk factors using statistical techniques such as linear regression. Such techniques assume a linear relationship between the predictor and dependent variables, thus oversimplifying and producing poor predictive power when the data has non-linear attributes (Byeon, 2015). It is therefore important to thoroughly identify and examine the risk factors associated with neonatal hearing loss through approaches that are able to determine more nuanced relationships between these variables.

New statistical advancements in machine learning techniques have been successfully implemented in various subdomains of audiology (Juhola et al., 2001; Anwar et al., 2010; Panchev et al., 2013; Shew et al., 2019; Zhao et al., 2019) and in the allocation of scarce resources across health systems (Price & Nicholson, 2014). The current study investigated a cohort of high-risk children with the aim of using predictive modelling to predict the occurrence of hearing loss using risk factors associated with neonatal hearing loss.

The risk factors present in the current study are widespread and consistent with those listed in the HPCSA (2018), JCIH (2019), and other related studies. Ototoxic medication, family history, and extracorporeal membrane oxygenation (ECMO) were the most frequently reported risk factors. Interestingly, maternal and/or infant HIV infection was among the top 10 frequently reported risk factors indicating that the profile of high-risk neonates may differ across contexts especially in developing contexts where the burden of disease (such as HIV/AIDS) may influence the way in which the risk factors interact with one another. Further analysis of these risk factors was conducted in order to establish the risk factors which were more likely to determine the occurrence of neonatal hearing loss.

Apart from hyperbilirubinemia, Table 2 clearly shows that all risk factors have been found to have statistically significant associations with hearing loss[2]. However, two risk factors namely, prematurity and mode of delivery, which are significantly associated with neonatal hearing loss in the current study have not yet been added to the list of risk factors listed by the HPCSA (2018) and the JCIH (2019). Notwithstanding, several studies have found associations between these prematurity and mode of delivery and neonatal hearing loss.

Hearing loss has been a commonly quoted consequence of prematurity (Marlow, Hunt, & Marlow, 2000; Wroblewska-Seniuk, Greczka, Dabrowski, Szyfter-Harris, & Mazela, 2017; Wang et al., 2018; Robertson et al., 2009). The pathophysiology of neonatal hearing loss in premature infants is complex and, as with other neurological structures, the period between 20 – 33 weeks is a critical period in gestation in which audiological function development occurs. (Marlow et al., 2000). While prematurity alone may not have a strong impact on hearing, it is commonly associated with other risk factors which, in combination, may predispose the infant to having a hearing loss. The risk factors which particularly occur alongside prematurity include exposure to ototoxic medication (i.e., aminoglycosides, C-reactive protein, loop diuretics) (Wroblewska-Seniuk et al., 2017; Wang et al., 2018), prolonged exposure to noise generated by the life-support machinery in NICU (Wroblewska-Seniuk et al., 2017), hyperbilirubinemia (Marlow et al., 2000; Wroblewska-Seniuk et al., 2017), hypoxia (Wroblewska-Seniuk et al., 2017), and the resultant prolonged oxygen use (Robertson et al., 2019).

Mode of delivery is an additional risk factor which been revealed to be a determining risk factor for neonatal hearing loss in the current study. Several studies examining the role of

[2] These associations are purely correlational and thus causality cannot be claimed.

birth type on infant hearing status has placed much focus on the differences in the effects that caesarean and natural birth have on newborn hearing screening outcomes (Oghan, Guvey, Topuz, Erdogan & Guvey, 2020; Xiao, Li, Xiao, Jiang, & Hu, 2015; Mylonas & Friese, 2015; Guven, 2019; Farahani, Mahrani, Seifrabiei, & Emadi, 2017; Smolkin et al., 2012; Olusanya & Solanke, 2009). The recurrent finding from these studies indicated that infants are more likely to fail their first hearing screening test due to the delayed absorption of middle ear fluid if they were born via caesarean delivery than via natural delivery (Oghan et al., 2020; Xiao et al., 2015; Mylonas & Friese, 2015; Guven, 2019; Farahani et al., 2017; Smolkin et al., 2012; Olusanya & Solanke, 2009). Furthermore, studies investigating the blood concentrations of the anaesthetics administered during the caesarean section procedure noted that the effects of the anaesthetics slowed down the rate of the neurotransmission of sound thus resulting in failed screening results (Diaz et al., 1977; Khoza-Shangase & Joubert, 2011). While these findings are important in understanding the high referral rates in infants born via caesarean birth, it does not necessarily mean that these infants will go on to develop a hearing loss. It can be inferred rather that the statistically significant association between caesarean delivery and hearing loss in the current study is a result of premature births occurring via caesarean delivery. A study conducted by Naidoo and Moodley (2009) on the increasing rates of caesarean sections in South Africa found that infants were being delivered via caesarean section from as early as 28 weeks for elective caesareans and 27 weeks for emergency caesareans. It can be deduced rather, that the prevalence of hearing loss in caesarean delivered babies are inversely related to the maturity of the baby rather than the procedure itself. On the other hand, the statistically significant association between natural birth and hearing loss could be due to the presence of other and/or unknown risk factors present in the cohort which, in combination, contribute significantly to the occurrence of hearing loss. However, the significant association between mode of delivery and hearing loss is purely

correlational therefore causality cannot be claimed (though this could be argued for infants born via caesarean section if they are being born prematurely, i.e., before the 37th gestational week).

The predictive relationship between high-risk factors and neonatal hearing loss has not been thoroughly investigated. In the South African context specifically, it is to the author's knowledge that there have been no studies conducted which aimed at developing prediction models using machine learning techniques for the prediction of hearing loss in high-risk neonates. Artificial intelligence (AI) and its subdomain machine learning have been slow in assimilating themselves into healthcare despite their potential to improve care, save lives, and cut costs especially in developing countries (Shew et al., 2019). Machine learning has enabled researchers to analyse large amounts of high-dimensional data, discover associations and patterns within this data, and ultimately build and apply models from this data to predict diagnostic results on unseen medical data based on what the model has learned. Previous studies in audiology have focused on the inferences related to the causal effects that risk factors have on the hearing thresholds of children and how a combination of these risk factors relate to one another (Beswick et al., 2013; Ohl et al., 2009; Bielecki et al., 2011). Machine learning, on the other hand, allows one to supply unanalysed audiological data into learned models and subsequently make accurate predictions about neonatal hearing loss. By analysing the function of risk factors associated with neonatal hearing loss, one may be able to indirectly identify important information about how these risk factors, independently and in combination with one another, effect the development of hearing loss in early life.

Two classification models for the prediction of neonatal hearing loss based on the high-risk factors associated with it has been presented. Both methods provided prediction accuracies between 88% and 92% with the logistic regression model performing slightly

better than the random forest model in terms of classification rate. The slight deviance in predictive accuracy can be explained by the premises on which each algorithm is based. Firstly, random forest models usually perform best when they are constructed using a small number of predicting variables (Kirasich, Smith, & Sadler, 2018). This is in contrast to logistic regression where, regardless of the number of predictors, the accuracy of the model continues to increase (Kirasich et al., 2018). The complexity of the random forest model, due to being constructed using several potential risk factors may have resulted in overfitting thus decreasing the model's predictive accuracy. However, the random forest algorithm with which the model was built corrects for this issue by using an ensamble of decision tress where the values on each tree originate from a random, independent sample (Kirasich et al., 2018). Moreover, the model was tuned during the training process to achieve the best possible prediction accuracy. Lastly, random forest models tend to underperform if the test data is far from the range of the testing data (Wickramanoyake, 2017). This means that although the model can accurately predict the same data it was trained on, it may underperform when applied to data where similar patterns and variations as the training set does not exist. On the contrary, logistic regression models has the ability to extrapolate data well outside the range of the training data as the logistic regression algorithm is constructed on arithmetic function (Wickramanoyake, 2017). The random forest model may have underperformed when compared to the logistic regression model as the variables present in the training dataset were not adequately presented in the testing dataset.

Furthermore, both models placed predictive importance on the same variables namely, prematurity, family history, cultural positive infections, and meningitis. Although this does not relate to the model's accuracy, it relates to the importance of each predictor making a prediction about neonatal hearing loss. While previous studies have examined the risk factors related to neonatal hearing loss, they have not specifically measured the predictive

importance of each factor. Nonetheless, prematurity, family history, cultural-positive infections and meningitis have been consistently reported amongst the most prevalent risk factors for hearing loss (Beswick et al., 2013; le Roux, Swanepoel, Louw, & Vinck, 2015; Beswick, Driscoll, Kei, & Glennon, 2012; Olusanya, 2011a; Khan et al., 2018; Martines et al., 2012; Kanji & Khoza-Shangase, 2019; Beswick, Driscoll, & Kei, 2012; Kuschke & Peer, 2018).

4.1. Contributions of the Current Study

The present study has demonstrated the use of machine learning techniques to predict the occurrence of hearing loss in high-risk neonates. Understanding the prevalence and predictive important of the risk factors associated with neonatal hearing loss is crucial in informing further exploration of these risk factors to provide a more comprehensible understanding of its effects on audiological functioning. This study has not only provided an understanding of the risk factors which may have the most effect on audiological functioning in a South African context, but also offers a novel method to identify infants who have a potential risk of late onset/progressive hearing loss. Prematurity, family history, cultural-positive infections, and meningitis have been found in the current study to have the most predictive power in determining neonatal hearing loss. These results have provided information that can be used as an initial step towards developing a high-risk registry for neonatal hearing loss (specifically for South African children). High-risk registries are important as it helps to alert medical practitioners to suspect the presence of a late onset/progressive hearing loss, which will in turn result in early identification and intervention. Results from the current study has shown the potential of using machine learning techniques to identify infants who are at increased risk for late onset/progressive hearing loss. Implementing these techniques in real time allows practitioners using these

models to identify at-risk neonates (especially those who reside in remote areas where some form of UNHS is not implemented) and refer them to audiological monitoring or targeted surveillance programmes. This will subsequently result in hearing identification and intervention services which will minimise the developmental effects of neonatal hearing loss.

4.2. Limitations of the Current Study

While implementing machine learning techniques into neonatal audiology offers an exciting new prospect, several limitations of the current study must be acknowledged. Firstly, the 'black-box' nature of machine learning algorithms can be difficult to interpret. While one has control over the data inputted into the model and is knowledgeable about what predictions one wishes to make (i.e., the output), there is a limited understanding of the process which occurs between the two as it is opaque (Shew et al., 2019; Price, 2018). As a result, the way in which the risk factor variables interact along with their independent effects on the outcome is unknown (Weng et al., 2017) and thus meaningful inferences about how the mode is built cannot be determined. Furthermore, apart from regression models, there are no confidence intervals or odds ratios equivalent for machine learning algorithms (Shew et al., 2019). This means that although predictor variables can be assessed to understand its contribution to the model, the extent of this contribution is unknown (Shew et al., 2019). In spite of the ongoing efforts to understand and unlock this 'black-box' such as improvements in data visualization methods, a definite solution has not yet been uncovered (Price et al., 2018).

Secondly, the current study investigated each risk factor in its entirely and did not take into consideration the specific indicators which each risk factor consisted of (e.g., cytomegalovirus and rubella were both measured as in-utero infection). This limits the

understanding of the effect that each individual sub-domain of a risk factor has on audiological functioning. Future studies should examine each medical condition which makes up the individual risk factor in order to obtain a more comprehensive understanding of individual effects on neonatal hearing loss. Moreover, due to the small frequency of certain risk factors, hyperbilirubinemia and neurodegenerative disorders could not be included in both machine learning models. A more widescale, comprehensive dataset should be used in order to successfully analyse these two risk factors.

Thirdly, while both datasets provide a comprehensive and holistic overview of infant hearing loss, the datasets vary in terms of healthcare status. The HI HOPES dataset includes a combination of infants in both private and public healthcare whereas the Netcare dataset only consists of infants who are in private healthcare. This may affect the accuracy of the model's prediction as certain risk factors may present more often in children from differing socioeconomic status groups. Socioeconomic status should be included in future machine learning models to determine whether it has an effect on the factors influencing neonatal hearing loss. Additionally, the data collected from Netcare is dependent on a data recording and management system whereby data may be incorrectly entered into the system due to human error. However, this challenge is limited as the data is entered by trained healthcare professionals, so the risk of error is limited.

Lastly, the dataset used in the current study was highly unbalanced with more hearing children being represented than deaf and/or hard-of-hearing children. Although this imbalance was addressed through oversampling, it is important to note that this is not the same as obtaining new data points. In essence, the data of deaf and hard-of-hearing neonates were being resampled. As such, the results may not be generalisable to the wider population.

Chapter 5: Conclusion

Machine learning has played a unique role in both general healthcare and in the field of audiology. Despite its limitations, machine learning has shown its potential role in predicting hearing outcomes. The current research study has contributed to existing research by examining the risk factors associated with neonatal hearing loss and constructing machine learning models to predict the occurrence of hearing loss in high-risk neonates. Results have extended and supported previous studies by providing new methodologies to the field of neonatal audiology as well as early childhood hearing and detection strategies.

References

Abdalla, F. M., & Omar, M. A. (2011). The role of the health system in the prevention of hearing loss among children in Sub-Saharan Africa. *Sudanese journal of paediatrics*, *11*(1), 8.

Abdulqader, Q. M. (2017). Applying the binary logistic regression analysis on the medical data. *Science Journal of University of Zakho*, *5*(4), 330-334.

Abidin, N. Z., Ismail, A. R., & Emran, N. A. (2018). Performance analysis of machine learning algorithms for missing value imputation. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *9*(6), 442-447.

Admiraal, R. J. C., & Huygen, P. L. M. (1999). Causes of hearing impairment in deaf pupils with a mental handicap. *International journal of pediatric otorhinolaryngology*, *51*(2), 101-108.

Ainsworth, M. S., & Bowlby, J. (1991). An ethological approach to personality development. *American psychologist*, *46*(4), 333.

Algood, C. L., Hong, J. S., Gourdine, R. M., & Williams, A. B. (2011). Maltreatment of children with developmental disabilities: An ecological systems analysis. *Children and Youth Services Review*, *33*(7), 1142-1148.

Alonso, S. G., de la Torre-Díez, I., Hamrioui, S., López-Coronado, M., Barreno, D. C., Nozaleda, L. M., & Franco, M. (2018). Data mining algorithms and techniques in mental health: A systematic review. *Journal of medical systems*, *42*(9), 1-15.

Ambili, K., & Afsar, P. (2016). A Prediction Model for Child Development Analysis using Naive Bayes

Amrit, C., Paauw, T., Aly, R., & Lavric, M. (2017). Identifying child abuse through text mining and machine learning. *Expert systems with applications*, *88*, 402-418.

and Decision Tree Fusion Technique – NB Tree. *International Research Journal of Engineering and Technology*, *3*(7), 402 -207.

Anwar, M. N., Oakes, M. P., Wermter, S., & Heinrich, S. (2010). Clustering audiology data. In *19th Machine Learning Conference of Belgium and The Netherlands*, Leuven, Belgium

Asiri, S. (2018). Machine Learning Classifiers. *Towards Data Science*. https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623

Bakhos, D., Marx, M., Villeneuve, A., Lescanne, E., Kim, S., & Robier, A. (2017). Electrophysiological exploration of hearing. *European annals of otorhinolaryngology, head and neck diseases*, *134*(5), 325-331.

Batista, G. E., & Monard, M. C. (2003). An analysis of four missing data treatment methods for supervised learning. *Applied artificial intelligence*, *17*(5-6), 519-533.

Benito-Orejas, J. I., Ramirez, B., Morais, D., Almaraz, A., & Fernández-Calvo, J. L. (2008). Comparison of two-step transient evoked otoacoustic emissions (TEOAE) and automated auditory brainstem response (AABR) for universal newborn hearing screening programs. *International journal of pediatric otorhinolaryngology*, *72*(8), 1193-1201.

Berninger, E., & Westling, B. (2011). Outcome of a universal newborn hearing-screening programme based on multiple transient-evoked otoacoustic emissions and clinical brainstem response audiometry. *Acta oto-laryngologica*, *131*(7), 728-739.

Beswick, R., Driscoll, C., & Kei, J. (2012). Monitoring for postnatal hearing loss using risk factors: a systematic literature review. *Ear and hearing*, *33*(6), 745-756.

Beswick, R., Driscoll, C., Kei, J., & Glennon, S. (2012). Recommendations for monitoring hearing in children using a risk factor registry. *Journal of Hearing Science*, *2*(3), 57-65.

Beswick, R., Driscoll, C., Kei, J., & Glennon, S. (2012). Targeted surveillance for postnatal hearing loss: A program evaluation. *International Journal of Pediatric Otorhinolaryngology*, *76*(7), 1046-1056.

Bezuidenhout, J. K. *Feasibility assessment of a universal newborn hearing screening programme at Rahima Moosa mother and child hospital* (Doctoral dissertation).

Bezuidenhout, J. K., Khoza-Shangase, K., De Maayer, T., & Strehlau, R. (2018). Universal newborn hearing screening in public healthcare in South Africa: Challenges to implementation. *South African Journal of Child Health*, *12*(4), 159-163.

Bhamjee, A., Le Roux, T., Schlemmer, K., Perold, J., Cass, N., Schroeder, K., ... & Vinck, B. (2019). Parent-perceived challenges related to the pediatric cochlear implantation process and support services received in South Africa. *International journal of pediatric otorhinolaryngology*, *126*, 109635.

Bielecki, I., Horbulewicz, A., & Wolan, T. (2012). Prevalence and risk factors for auditory neuropathy spectrum disorder in a screened newborn population at risk for hearing loss. *International journal of pediatric otorhinolaryngology*, *76*(11), 1668-1670.

Brofenbrenner, U. (1994). Ecological models of human development. *International encyclopedia of education*, *3*(2), 1643-1647.

Bronfenbrenner, U. (1974). Developmental research, public policy, and the ecology of childhood. *Child development*, *45*(1), 1-5.

Bruyere, E., & Garbarino, J. (). The ecological perspective on the human rights of children. In S. Kamerman, S., Phipps., & A. Ben-Arieh. (Eds.). (2009). *From child welfare to*

child well-being: An international perspective on knowledge in the service of policy making (Vol. 1). Springer Science & Business Media.

Burnash, D. L., Rothman-Marshall, G., & Schenkel, L. S. (2010, November). Child maltreatment in deaf college students: An analysis of the prevalence, characteristics, and clinical outcomes. In *Poster presented at the Association for Behavioral and Cognitive Therapies (ABCT) Annual Convention (Child Maltreatment SIG) San Francisco, CA*.

Byeon, H. (2015). A prediction model for mild cognitive impairment using random forests. *Int J Adv Comput Sci App*, *6*(12), 8.

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, *110*(1-12), 24.

Cicchetti, D., Toth, S. L., & Maughan, A. (2000). An ecological–transactional model of child maltreatment.

Ciorba, A., Hatzopoulos, S., Busi, M., Guerrini, P., Petruccelli, J., & Martini, A. (2008). The universal newborn hearing screening program at the University Hospital of Ferrara: focus on costs and software solutions. *International journal of pediatric otorhinolaryngology*, *72*(6), 807-816.

Colgan, S., Gold, L., Wirth, K., Ching, T., Poulakis, Z., Rickards, F., & Wake, M. (2012). The cost-effectiveness of universal newborn screening for bilateral permanent congenital hearing impairment: systematic review. *Academic pediatrics*, *12*(3), 171-180.

Cone-Wesson, B., Vohr, B. R., Sininger, Y. S., Widen, J. E., Folsom, R. C., Gorga, M. P., & Norton, S. J. (2000). Identification of neonatal hearing impairment: infants with hearing loss. *Ear and hearing*, *21*(5), 488-507.

Copley, G. J., & Friderichs, N. B. (2010). An approach to hearing loss in children. *South African Family Practice*, *52*(1), 34-39.

Creswell, J., 2005. *Research design*. Los Angeles: Sage Publications Inc.

Crouch, J. L., & Behl, L. E. (2001). Relationships among parental beliefs in corporal punishment, reported stress, and physical child abuse potential. *Child abuse & neglect*, *25*(3), 413-419.

Cunningham, R. F. (2011). Otoacoustic emissions: Beyond newborn hearing screening. *Audiology Online*. https://www.audiologyonline.com/articles/otoacoustic-emissions-beyond-newborn-hearing-838#:~:text=An%20otoacoustic%20emission%20(OAE)%20is,evoked%20by%20an%20auditory%20stimulus.&text=Today%2C%20OAEs%20are%20used%20commonly,not%20volunteer%20reliable%20behavioral%20responses.

Davids, R. S., & De Jager, M. (2018). Experiences of hearing parents following their child's hearing loss diagnosis. *Southern African Journal of Social Work and Social Development*, *30*(2), 1-18.

de Andrade, V., & Ross, E. (2005). Beliefs and practices of Black South African traditional healers regarding hearing impairment: Creencias y prácticas de los curanderos negros sudafricanos en torno a la hipoacusia. *International Journal of Audiology*, *44*(9), 489-499.

Dereköy, F. S. (2000). Etiology of deafness in Afyon school for the deaf in Turkey. *International journal of pediatric otorhinolaryngology*, *55*(2), 125-131.

Diaz, M., Graff, M., Hiatt, I. M., Friedman, S., Ostfeld, B., & Hegyi, T. (1988). Prenatal lidocaine and the auditory evoked responses in term infants. *American Journal of Diseases of Children*, *142*(2), 160-161.

Ding, Y., & Simonoff, J. S. (2006). An Investigation of Missing Data Methods for Classiffcation Trees. *Statistics Working Papers Series, Vol*.

Donges, N. (2020). A complete guide to the random forest algorithm. *Built in.* https://builtin.com/data-science/random-forest-algorithm

Durieux-Smith, A., & Whittingham, J. (2000). The Rationale for Neonatal Hearing Screening. *The Journal of Speech-Language Pathology and Audiology*, *24*(2).

Egeli, E., Çiçekci, G., Silan, F., Öztürk, Ö., Harputluoğlu, U., Onur, A., ... & Yildiz, A. (2003). Etiology of deafness at the Yeditepe school for the deaf in Istanbul. *International journal of pediatric otorhinolaryngology*, *67*(5), 467-471.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, *5*(1), 1-4.

Farahani, F., Nahrani, M. H., Seifrabiei, M. A., & Emadi, M. (2017). The effect of mode of delivery and hospital type on newborn hearing screening results using otoacoustic emissions: based on screening age. *Indian Journal of Otolaryngology and Head & Neck Surgery*, *69*(1), 1-5.

Fawcett, T., & Hardin, D. (2017). Machine Learning vs. Statistics: The Texas Death Match of Data Science. *KDnuggets*. https://www.kdnuggets.com/2017/08/machine-learning-vs-statistics.html

Findlen, U. M., Hounam, G. M., Alexy, E., & Adunka, O. F. (2019). Early hearing detection and intervention: timely diagnosis, timely management. *Ear and hearing*, *40*(3), 651-658.

Flannelly, L. T., Flannelly, K. J., & Jankowski, K. R. (2014). Independent, dependent, and other variables in healthcare and chaplaincy research. *Journal of health care chaplaincy*, *20*(4), 161-170.

Fleischman, A. R., Oinuma, M., & Clark, S. L. (2010). Rethinking the definition of "term pregnancy". *Obstetrics & Gynecology*, *116*(1), 136-139.

Gabriel, M. M., Geyer, L., McHugh, C., Thapa, J., Glynn, F., Walshe, P., ... & Viani, L. (2020). Impact of Universal Newborn Hearing Screening on cochlear implanted children in Ireland. *International journal of pediatric otorhinolaryngology*, *133*, 109975.

Gardiner, S. A., Laing, N., Mall, S., & Wonkam, A. (2019). Perceptions of parents of children with hearing loss of genetic origin in South Africa. *Journal of community genetics*, *10*(3), 325-333.

Geisser, S. (1993). *Predictive inference* (Vol. 55). CRC press.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., & International Child Development Steering Group. (2007). Developmental potential in the first 5 years for children in developing countries. *The lancet*, *369*(9555), 60-70.

Gruss, I., Berlin, M., Greenstein, T., Yagil, Y., & Beiser, M. (2007). Etiologies of hearing impairment among infants and toddlers: 1986–1987 versus 2001. *International journal of pediatric otorhinolaryngology*, *71*(10), 1585-1589.

Güven, S. G. (2019). The effect of mode of delivery on newborn hearing screening results. *Turkish archives of otorhinolaryngology*, *57*(1), 19.

Hall III, J. W. (2015). Combined OAE and AABR Approach for Newborn Hearing Screening. *Audiology Online. https://www.audiologyonline.com/articles/combined-oae-and-aabr-approach-15543*

Han, J., Kamber, M., & Pei, J. (2011). Data mining concepts and techniques third edition. *The Morgan Kaufmann Series in Data Management Systems*, *5*(4), 83-124.

Harper, P. R. (2005). A review and comparison of classification algorithms for medical decision making. *Health Policy*, *71*(3), 315-331.

Harvey, M. A., & Dym, B. (1987). An ecological view of deafness. *Family Systems Medicine*, *5*(1), 52.

Health Professions Council of South Africa. (2018). Early Hearing Detection and Intervention Programmes in South Africa, Position Statement Year 2018. *Health Professions Council of South Africa.* https://www.hpcsa.co.za/Uploads/SLH/Guidelines%20for%20Early_Hearing_Detection_and_Intervention_(EHDI)_2018.pdf

Heaton, J. B., Polson, N. G., & Witte, J. H. (2016). Deep learning in finance. *arXiv preprint arXiv:1602.06561*.

Horn, D. L., Pisoni, D. B., & Miyamoto, R. T. (2006). Divergence of fine and gross motor skills in prelingually deaf children: implications for cochlear implantation. *The Laryngoscope*, *116*(8), 1500-1506.

Husain, W., Xin, L. K., & Jothi, N. (2016, August). Predicting Generalized Anxiety Disorder among women using random forest approach. In *2016 3rd International Conference on Computer and Information Sciences (ICCOINS)* (pp. 37-42). IEEE.

Hyde, M. L. (2005). Newborn hearing screening programs: Overview. *Journal of Otolaryngology*, *34*(2), S70.

Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and cognition*, *7*(2), 265-290.

Imam, S. S., El-Farrash, R. A., Taha, H. M., & Bishoy, H. E. (2013). Targeted versus universal neonatal hearing screening in a single Egyptian center. *International Scholarly Research Notices*, *2013*.

Jacob, J., Lehne, M., Mischker, A., Klinger, N., Zickermann, C., & Walker, J. (2017). Cost effects of preterm birth: a comparison of health care costs associated with early preterm, late preterm, and full-term birth in the first 3 years after birth. *The European Journal of Health Economics, 18*(8), 1041-1046.

Jacob-Corteletti, L. C. B., Araújo, E. S., Duarte, J. L., Zucki, F., & Alvarenga, K. D. F. (2018). Acoustic Reflex Testing in Neonatal Hearing Screening and Subsequent Audiological Evaluation. *Journal of Speech, Language, and Hearing Research*, *61*(7), 1784-1793.

Jha, A. K., DesRoches, C. M., Campbell, E. G., Donelan, K., Rao, S. R., Ferris, T. G., ... & Blumenthal, D. (2009). Use of electronic health records in US hospitals. *New England Journal of Medicine*, *360*(16), 1628-1638.

Johnson, C. P., & Blasco, P. A. (1997). Infant growth and development. *Pediatr Rev*, *18*(7), 224-242.

Joint Committee on Infant Hearing. (2019). Year 2019 Position Statement: Principles and Guidelines for Early Hearing Detection and Intervention Programs. *The Journal of Early Hearing Detection and Intervention.* https://digitalcommons.usu.edu/jehdi/vol4/iss2/1

Juhola, M., Viikki, K., Laurikkala, J., Auramo, Y., Kentala, E., & Pyykkö, I. (2001). Application of artificial intelligence in audiology. *Scandinavian Audiology*, *30*(1), 97-99.

Juhola, M., Viikki, K., Laurikkala, J., Auramo, Y., Kentala, E., & Pyykkö, I. (2001). Application of artificial intelligence in audiology. *Scandinavian Audiology*, *30*(1), 97-99.

Kamal, N. (2013). Newborn hearing screening: opportunities and challenges. *Egyptian Journal of Ear, Nose, Throat and Allied Sciences*, *14*(2), 55-58.

Kanji, A., & Khoza-Shangase, K. (2012). The occurrence of high-risk factors for hearing loss in verylow-birth-weight neonates: A retrospective exploratory study of targeted hearing screening. *South African Journal of Communication Disorders*, *59*(1), 3-7.

Kanji, A., & Khoza-Shangase, K. (2019). Risk factors for hearing impairment in neonates in South Africa: scoping the context for newborn hearing screening planning. *The Journal of Maternal-Fetal & Neonatal Medicine*, 1-10.

Kanji, A., Khoza-Shangase, K., & Ballot, D. (2010). Hearing screening follow-up return rate in a very low birth weight project: A retrospective record review. *South African Journal of Child Health*, *4*(4), 95-99.

Kapitanoff, S. H., Lutzker, J. R., & Bigelow, K. M. (2000). Cultural issues in. *Aggression and Violent Behavior*, *5*(3), 227-244.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*, *15*, 104-116.

Kemp, D. T. (2002). Otoacoustic emissions, their origin in cochlear function, and use. *British medical bulletin*, *63*(1), 223-241.

Kennedy, C. R. (2000). Neonatal screening for hearing impairment. *Archives of disease in childhood*, *83*(5), 377-383.

Khan, N. B., Joseph, L., & Adhikari, M. (2018). The hearing screening experiences and practices of primary health care nurses: Indications for referral based on high-risk factors and community views about hearing loss. *African Journal of Primary Health Care & Family Medicine*, *10*(1), 1-11.

Khoza-Shangase, K. (2019). Early hearing detection and intervention in South Africa: Exploring factors compromising service delivery as expressed by caregivers. *International journal of pediatric otorhinolaryngology*, *118*, 73-78.

Khoza-Shangase, K. (2019). Early hearing detection and intervention in South Africa: Exploring factors compromising service delivery as expressed by caregivers. *International journal of pediatric otorhinolaryngology*, *118*, 73-78.

Khoza-Shangase, K., & Joubert, K. (2011). The influence of epidural anesthesia on new-born hearing screening: A pilot study. *Journal of Pharmacy and Bioallied Sciences*, *3*(1), 135.

Kim, H., & Yates, S. (2003). Missing value algorithms in decision trees. In *Statistical data mining and knowledge discovery* (pp. 155-172). CRC Press.

Kim, S. H., Choi, H. S., Han, Y. E., & Choi, B. Y. (2016). Diverse etiologies manifesting auditory neuropathy characteristics from infants with profound hearing loss and clinical implications. *International journal of pediatric otorhinolaryngology*, *86*, 63-67.

Kirasich, K., Smith, T., & Sadler, B. (2018). Random Forest vs logistic regression: binary classification for heterogeneous datasets. *SMU Data Science Review*, *1*(3), 9.

Knutson, J. F., Johnson, C. R., & Sullivan, P. M. (2004). Disciplinary choices of mothers of deaf children and mothers of normally hearing children. *Child Abuse & Neglect*, *28*(9), 925-937.

Kose, I., Gokturk, M., & Kilic, K. (2015). An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. *Applied Soft Computing*, *36*, 283-299.

Kossewska, J. (2016). *Studies on deafness in an ecological system context*. Wydawnictwo JAK.

Kotby, M. N., Tawfik, S., Aziz, A., & Taha, H. (2008). Public health impact of hearing impairment and disability. Folia Phoniatrica et Logopaedica, 60(2), 58-63.

Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised leaning. *International Journal of Computer Science*, *1*(2), 111-117.

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, *30*(1), 25-36.

Kube, A. R., Das, S., & Fowler, P. J. (2019). Fair and E cient Allocation of Scarce Resources Based on Predicted Outcomes: Implications for Homeless Service Delivery.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). New York: Springer.

Kuschke, S., Goncalves, N., & Peer, S. (2018). Hearing outcomes in children with meningitis at Red Cross War Memorial Children's Hospital, Cape Town, South Africa: A silent crisis. *South African Medical Journal*, *108*(11), 944-946.

Kuschke, S., Goncalves, N., & Peer, S. (2018). Hearing outcomes in children with meningitis at Red Cross War Memorial Children's Hospital, Cape Town, South Africa: A silent crisis. *South African Medical Journal*, *108*(11), 944-946.

Kushalnagar, P., Topolski, T. D., Schick, B., Edwards, T. C., Skalicky, A. M., & Patrick, D. L. (2011). Mode of communication, perceived level of understanding, and perceived quality of life in youth who are deaf or hard of hearing. *Journal of deaf studies and deaf education*, *16*(4), 512-523.

Kvam, M. H. (2004). Sexual abuse of deaf children. A retrospective analysis of the prevalence and characteristics of childhood sexual abuse among deaf adults in Norway. *Child abuse & neglect*, *28*(3), 241-251.

Lansford, J. E., Dodge, K. A., Pettit, G. S., Bates, J. E., Crozier, J., & Kaplow, J. (2002). A 12-year prospective study of the long-term effects of early child physical maltreatment on psychological, behavioral, and academic problems in adolescence. *Archives of pediatrics & adolescent medicine*, *156*(8), 824-830.

Le Roux, T., Swanepoel, D. W., Louw, A., Vinck, B., & Tshifularo, M. (2015). Profound childhood hearing loss in a South Africa cohort: Risk profile, diagnosis and age of intervention. *International Journal of Pediatric Otorhinolaryngology*, *79*(1), 8-14.

Levy, K. N., & Blatt, S. J. (1999). Attachment theory and psychoanalysis: Further differentiation within insecure attachment patterns. *Psychoanalytic Inquiry*, *19*(4), 541-575.

Liu, C., Wang, A., Huang, Y., Zhang, Y., Ding, H., Wu, J., ... & Yin, A. (2019). Development of a community-based hearing loss prevention and control service model in Guangdong, China. *BMC public health*, *19*(1), 1-7.

Liu, W. Z., White, A. P., Thompson, S. G., & Bramer, M. A. (1997). Techniques for dealing with missing values in classification. In *International Symposium on Intelligent Data Analysis* (pp. 527-536). Springer, Berlin, Heidelberg.

Livingston, F. (2005). Implementation of Breiman's random forest machine learning algorithm. *ECE591Q Machine Learning Journal Paper*, 1-13.

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: A Package for Binary Imbalanced Learning. *R journal*, *6*(1).

Luo, J., Wu, M., Gopukumar, D., & Zhao, Y. (2016). Big data application in biomedical research and health care: a literature review. *Biomedical informatics insights, 8*, BII-S31559.

Mackenzie, I., & Smith, A. (2009). Deafness—the neglected and hidden disability. *Annals of Tropical Medicine & Parasitology*, *103*(7), 565-571.

MAICO Diagnostics. (2009). easyScreen – ABR + OAE Combined in One Hearing Screening Device. *MAICO. https://www.maico-diagnostics.com/products/oae/easyscreen*

Marlow, E. S., Hunt, L. P., & Marlow, N. (2000). Sensorineural hearing loss and prematurity. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *82*(2), F141-F144.

Martines, F., Salvago, P., Bentivegna, D., Bartolone, A., Dispenza, F., & Martines, E. (2012). Audiologic profile of infants at risk: experience of a Western Sicily tertiary care centre. *International journal of pediatric otorhinolaryngology*, *76*(9), 1285-1291.

Martínez-Cruz, C. F., Poblano, A., & Fernández-Carrocera, L. A. (2008). Risk factors associated with sensorineural hearing loss in infants at the neonatal intensive care unit: 15-year experience at the National Institute of Perinatology (Mexico City). *Archives of medical research*, *39*(7), 686-694.

McCreery, R. (2013). Otoacoustic emissions: Beyond “pass” and “refer”. *The Hearing Journal*, *66*(9), 14-5.

McPherson, B., & Swart, S. M. (1997). Childhood hearing loss in sub-Saharan Africa: a review and recommendations. *International journal of pediatric otorhinolaryngology*, *40*(1), 1-18.

Melton, G. B. (2008). Beyond balancing: Toward an integrated approach to children's rights. *Journal of Social Issues*, *64*(4), 903-920.

Moeller, M. P. (2000). Early intervention and language development in children who are deaf and hard of hearing. *Pediatrics*, *106*(3), e43-e43.

Moodley, S. (2012). The role of the neonatal nurse in early hearing detection and intervention in South Africa. *Professional Nursing Today*, *16*(4).

Moodley, S. (2016). Paediatric diagnostic audiology testing in South Africa. *International journal of pediatric otorhinolaryngology*, *82*, 1-7.

Moodley, S., & Storbeck, C. (2017). Data management for early hearing detection and intervention in South Africa. *SA Journal of Information Management*, *19*(1), 6.

Morgan, A., & Vernon, M. (1994). A guide to the diagnosis of learning disabilities in deaf and hard-of-hearing children and adults. *American Annals of the Deaf*, 358-370.

Mylonas, I., & Friese, K. (2015). Indications for and risks of elective cesarean section. *Deutsches Ärzteblatt International*, *112*(29-30), 489.

Naidoo, N., & Moodley, J. (2009). Rising rates of Caesarean sections: an audit of Caesarean sections in a specialist private practice. *South African family practice*, *51*(3).

Nazir, T., Gupta, S., Mir, G. M., Jamwal, A., Kalsotra, P., & Singh, K. P. (2016). Evaluation of otoacoustic emissions and auditory brainstem responses for hearing screening of high risk infants. *Indian Journal of Otology*, *22*(4), 221.

Netcare. (2019). Netcare launches first national newborn hearing screening programme in South Africa. *Netcare Hospital Group*. https://www.netcare.co.za/News-Hub/Articles/netcare-launches-first-national-newborn-hearing-screening-programme-in-south-africa

Neumann, K., & Indermark, A. (2012). Validation of a new TEOAE-AABR device for newborn hearing screening. *International journal of audiology*, *51*(8), 570-575.

Northern, J. L. (2015). Marion downs: her life and legacy (1914–2014). *International Journal of Audiology*, *54*(8), 289-290.

NSW Department of Health. (2011). NSW Statewide Infant Screening - Hearing (SWIS-H) program. *NSW Government – Health*. https://www.health.nsw.gov.au/kidsfamilies/MCFhealth/child/Pages/hearing-services.aspx

Núñez-Batalla, F., Jáudenes-Casaubón, C., Sequí-Canet, J. M., Vivanco-Allende, A., & Zubicaray-Ugarteche, J. (2016). 2014 CODEPEH recommendations: early detection of late onset deafness, audiological diagnosis, hearing aid fitting and early intervention. *Acta Otorrinolaringologica (English Edition)*, *67*(1), 45-53.

Oghan, F., Guvey, A., Topuz, M. F., Erdogan, O., & Guvey, H. (2020). Effects of vaginal birth versus caesarean section on hearing screening results in a large series from the Aegean region. *International journal of audiology*, *59*(4), 310-315.

Ohl, C., Dornier, L., Czajka, C., Chobaut, J. C., & Tavernier, L. (2009). Newborn hearing screening on infants at risk. *International journal of pediatric otorhinolaryngology*, *73*(12), 1691-1695.

Olusanya, B. O. (2008). Priorities for early hearing detection and intervention in sub-Saharan Africa. *International journal of audiology*, *47*(sup1), S3-S13.

Olusanya, B. O. (2011a). Priorities for early childhood development in low-income countries. *Journal of Developmental & Behavioral Pediatrics*, *32*(6), 476-481.

Olusanya, B. O. (2011b). Highlights of the new WHO report on newborn and infant hearing screening and implications for developing countries. *International journal of pediatric otorhinolaryngology*, *75*(6), 745-748.

Olusanya, B. O., & Solanke, O. A. (2009). Predictors of term stillbirths in an inner-city maternity hospital in Lagos, Nigeria. *Acta obstetricia et gynecologica Scandinavica*, *88*(11), 1243-1251.

Olusanya, B. O., Luxon, L. M., & Wirz, S. L. (2005). Screening for early childhood hearing loss in Nigeria. *Journal of medical screening*, *12*(3).

Olusanya, B. O., Neumann, K. J., & Saunders, J. E. (2014). The global burden of disabling hearing impairment: a call to action. *Bulletin of the World Health Organization*, *92*, 367-373.

Panchev, C., Anwar, M. N., & Oakes, M. (2013). Hearing aid classification based on audiology data. In *International Conference on Artificial Neural Networks* (pp. 375-380). Springer, Berlin, Heidelberg.

Price, W., & Nicholson, I. I. (2014). Black-box medicine. *Harv. JL & Tech.*, *28*, 419.

Riga, M., Psarommatis, I., Lyra, C. H., Douniadakis, D., Tsakanikos, M., Neou, P., & Apostolopoulos, N. (2005). Etiological diagnosis of bilateral, sensorineural hearing impairment in a pediatric Greek population. *International journal of pediatric otorhinolaryngology*, *69*(4), 449-455.

Robertson, C. M., Howarth, T. M., Bork, D. L., & Dinu, I. A. (2009). Permanent bilateral sensory and neural hearing loss of children after neonatal intensive care because of extreme prematurity: a thirty-year study. *Pediatrics*, *123*(5), e797-e807.

RStudio. (2020). R. *RStudio*. https://www.rstudio.com/products/rstudio/

Saar-Tsechansky, M., & Provost, F. (2007). Handling missing values when applying classification models.

Sarica, A., Cerasa, A., & Quattrone, A. (2017). Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. *Frontiers in aging neuroscience*, *9*, 329.

Schenkel, L. S., Rothman-Marshall, G., Schlehofer, D. A., Towne, T. L., Burnash, D. L., & Priddy, B. M. (2014). Child maltreatment and trauma exposure among deaf and hard of hearing young adults. *Child abuse & neglect*, *38*(10), 1581-1589.

Schmidt, D., Niemann, M., & von Trzebiatowski, G. L. (2015). The Handling of Missing Values in Medical Domains with Respect to Pattern Mining Algorithms. In *CS&P* (pp. 147-154).

Serra, A., Galdi, P., & Tagliaferri, R. (2018). Machine learning for bioinformatics and neuroimaging. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(5), e1248.

Shew, M., New, J., Wichova, H., Koestler, D. C., & Staecker, H. (2019). Using machine learning to predict sensorineural hearing loss based on perilymph micro RNA expression profile. *Scientific reports*, *9*(1), 1-11.

SKI-HI Institute. (2004). SKI-HI Curriculum. Logan, UT: Hope, Inc.

Smolkin, T., Mick, O., Dabbah, M., Blazer, S., Grakovsky, G., Gabay, N., ... & Makhoul, I. R. (2012). Birth by cesarean delivery and failure on first otoacoustic emissions hearing test. *Pediatrics*, *130*(1), e95-e100.

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia medica*, *24*(1), 12-18.

Stewart, M. (2019). The Actual Difference Between Statistics and Machine Learning. *Towards Data Science*. https://towardsdatascience.com/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3

Storbeck, C., & Calvert-Evans, J. (2008). Towards integrated practices in early detection of and intervention for deaf and hard of hearing children. *American annals of the deaf*, *153*(3), 314-321.

Storbeck, C., & Moodley, S. (2010). ECD policies in South Africa What about children with disabilities?. *Journal of African Studies and Development*, *3*(1), 1-8.

Storbeck, C., & Moodley, S. (2011). ECD policies in South Africa What about children with disabilities?. *Journal of African Studies and Development*, *3*(1), 1-8.

Störbeck, C., & Young, A. (2016). The HI HOPES data set of deaf children under the age of 6 in South Africa: Maternal suspicion, age of identification and newborn hearing screening. *BMC pediatrics*, *16*(1), 1-10.

Swanepoel, D. W. (2009). Early detection of infant hearing loss in South Africa. *SAMJ: South African Medical Journal*, *99*(3), 158-159.

Swanepoel, D. W., Johl, L., & Pienaar, D. (2013). Childhood hearing loss and risk profile in a South African population. *International journal of pediatric otorhinolaryngology*, *77*(3), 394-398.

Swanepoel, D., Störbeck, C., & Friedland, P. (2009). Early hearing detection and intervention in South Africa. *International journal of pediatric otorhinolaryngology*, *73*(6), 783-786.

Swanepoel, H., & De Beer, F. (2012). *Community development: Breaking the cycle of poverty*. Juta and Company Ltd.

Tripathy, J. P. (2013). Secondary data analysis: Ethical issues and challenges. *Iranian journal of public health*, *42*(12), 1478.

Tucci, D. L., Merson, M. H., & Wilson, B. S. (2010). A summary of the literature on global hearing impairment: current status and priorities for action. *Otology & Neurotology*, *31*(1), 31-41.

Van der Heide, E. M. M., Veerkamp, R. F., Van Pelt, M. L., Kamphuis, C., Athanasiadis, I., & Ducro, B. J. (2019). Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle. *Journal of dairy science*, *102*(10), 9409-9421.

Van Dyk, M., Swanepoel, D. W., & Hall III, J. W. (2015). Outcomes with OAE and AABR screening in the first 48 h—Implications for newborn hearing screening in developing countries. *International journal of pediatric otorhinolaryngology*, *79*(7), 1034-1040.

Vélez-Agosto, N. M., Soto-Crespo, J. G., Vizcarrondo-Oppenheimer, M., Vega-Molina, S., & García Coll, C. (2017). Bronfenbrenner's bioecological theory revision: Moving culture from the macro into the micro. *Perspectives on Psychological Science*, *12*(5), 900-910.

Visser, M., & Moleko, A. G. (Eds.). (2012). *Community Psychology in South Africa*. Van Schaik Publishers.

Vos, B., Senterre, C., Lagasse, R., & Levêque, A. (2015). Newborn hearing screening programme in Belgium: a consensus recommendation on risk factors. *BMC pediatrics*, *15*(1), 1-14.

Wallace, B. C., Noel-Storr, A., Marshall, I. J., Cohen, A. M., Smalheiser, N. R., & Thomas, J. (2017). Identifying reports of randomized controlled trials (RCTs) via a hybrid

machine learning and crowdsourcing approach. *Journal of the American Medical Informatics Association*, *24*(6), 1165-1168.

Wang, L. A., Smith, P. B., Laughon, M., Goldberg, R. N., Ku, L. C., Zimmerman, K. O., ... & Greenberg, R. G. (2018). Prolonged furosemide exposure and risk of abnormal newborn hearing screen in premature infants. *Early human development*, *125*, 26-30.

Ware, A. (2018). Can Artificial Intelligence Alleviate Resource Scarcity?. *Inquiry Journal*.

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data?. *PloS one*, *12*(4), e0174944.

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data?. *PloS one*, *12*(4), e0174944.

Wickramanayake, B. (2017). Random Forest vs Logistic Regression. *Medium*. https://medium.com/@bemali_61284/random-forest-vs-logistic-regression-16c0c8e2484c

Wiens, J., & Shenoy, E. S. (2018). Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. *Clinical Infectious Diseases*, *66*(1), 149-153.

World Health Organisation. (2020, March 2). Deafness and Hearing Loss. *World Health Organisation*. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss#:~:text=Over%205%25%20of%20the%20world's,will%20have%20disabling%20hearing%20loss.

Wroblewska-Seniuk, K., Greczka, G., Dabrowski, P., Szyfter-Harris, J., & Mazela, J. (2017). Hearing impairment in premature newborns—Analysis based on the national hearing screening database in Poland. *PloS one*, *12*(9), e0184359.

Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J. F., & Hua, L. (2012). Data mining in healthcare and biomedicine: a survey of the literature. *Journal of medical systems*, *36*(4), 2431-2448.

Yoshikawa, S., Ikeda, K., Kudo, T., & Kobayashi, T. (2004). The effects of hypoxia, premature birth, infection, ototoxic drugs, circulatory system and congenital disease on neonatal hearing loss. *Auris Nasus Larynx*, *31*(4), 361-368.

Yoshinaga-Itano, C. (2004). Levels of evidence: universal newborn hearing screening (UNHS) and early hearing detection and intervention systems (EHDI). *Journal of communication disorders*, *37*(5), 451-465.

Yoshinaga-Itano, C., & Apuzzo, M. R. L. (1998b). Identification of hearing loss after age 18 months is not early enough. *American annals of the deaf*, 380-387.

Yoshinaga-Itano, C., Sedey, A. L., Coulter, D. K., & Mehl, A. L. (1998a). Language of early-and later-identified children with hearing loss. *Pediatrics*, *102*(5), 1161-1171.

You, T. (2019). Understanding Random Forest. *Towards Data Science*. https://towardsdatascience.com/understanding-random-forest-58381e0602d2

Zhang, S., Qin, Z., Ling, C. X., & Sheng, S. (2005). " Missing is useful": Missing values in cost-sensitive decision trees. *IEEE transactions on knowledge and data engineering*, *17*(12), 1689-1693.

Zhao, Y., Li, J., Zhang, M., Lu, Y., Xie, H., Tian, Y., & Qiu, W. (2019). Machine learning models for the hearing impairment prediction in workers exposed to complex industrial noise: a pilot study. *Ear and hearing*, *40*(3), 690.

Zhu, B., Gao, Z., Zhao, J., & vanden Broucke, S. K. (2019). IRIC: An R library for binary imbalanced classification. *SoftwareX*, *10*, 100341.

Appendix A

Plagiarism Declaration

University of the Witwatersrand, Johannesburg

School of Social Science

SENATE PLAGIARISM POLICY

Declaration by Students

I, Safiyyah Ismail, (Student number: 713549) am a student registered for MA eScience in the year 2020. I hereby declare the following:

- **I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.**
- **I confirm that ALL the work submitted for assessment for the above course is my own unaided work except where I have explicitly indicated otherwise.**
- **I have followed the required conventions in referencing the thoughts and ideas of others.**
- **I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this in not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.**

Signature:  **Date: 31st March 2021**

Appendix B

Ethical Approval

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG	HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research & Post Graduate Affairs)

TO: Ms S Ismail
School of Social Sciences
Centre for Deaf Studies
University

E-mail: 713549@students.wits.ac.za

CC: Supervisor: Dr S Moodley and Professor R Alence
<Selvarani.Moodley@wits.ac.za>
and <HREC-Medical.ResearchOffice@wits.ac.za>

FROM: Iain Burns
Human Research Ethics Committee (Medical)
Tel: 011 717 1252

E-mail: Iain.Burns@wits.ac.za

DATE: 2020/07/22

REF: *R14/49*

PROTOCOL NO: **M200516** *(This is your ethics application study reference number. Please quote this reference number in all correspondence relating to this study)*

PROJECT TITLE: *Detecting hearing loss in high-risk neonates using a machine learning algorithm*

Please find attached the Clearance Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to the Government funding of the University.

MSWorks2000/Iain0007/Clearscan.wps

Appendix C

Supervisor-Supervisee Contract

16. STATEMENT OF PRINCIPLES FOR POSTGRADUATE SUPERVISION

IN A CONTEXT OF ACADEMIC FREEDOM AND WITHIN A FRAMEWORK OF INDIVIDUAL AUTONOMY AND THE PURSUIT OF KNOWLEDGE THIS AGREEMENT IS WRITTEN IN THE BELIEF THAT THERE IS A RECIPROCAL RELATIONSHIP AND MUTUAL ACCOUNTABILITY BETWEEN SUPERVISOR AND STUDENT.

THE SUPERVISOR AND THE STUDENT:

1. Will establish agreed roles and clear processes to be maintained by both parties. In the case of joint supervision everyone's role needs to be clarified.
2. Will meet regularly and as frequently as is reasonable to ensure steady progress towards the completion of the proposal, research report, or dissertation or thesis. This time varies but the normal minimum requirement for face-to- face contact, spread across each year of registration is: 10 contact hours for an Honours project, 15 contact hours for a Masters by research report and 24 contact hours for a Masters by dissertation and a PhD.
3. Will keep appointments, be punctual and respond timeously to messages.
4. Will keep one another informed of any planned vacations or absences as well as changes in his or her personal circumstances that might impact on the work schedule. Unplanned absences or delays should be discussed as soon as possible and arrangements should be made to catch up lost time.
5. Will ensure that research on animal or human subjects is conducted according to the procedures and the requirements of the relevant University Ethics committee.
6. Will together complete progress reports on the research project, as requested by each Faculty Graduate Studies Committee.

THE SUPERVISOR:

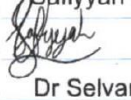
1. Undertakes to provide guidance for the student's research project in relation to the design and scope of the project, the relevant literature and information sources, research methods and techniques and methods of data analysis.
2. Has a responsibility to be accessible to the student.
3. Will be prepared for meetings with the student. This includes being up-to-date on the latest work in his/her area of expertise.
4. Will expect written work as jointly agreed, and will return that work with constructive criticism within a timeframe (a suggestion of 2-4 weeks) jointly agreed at the outset of the research.
5. Will provide advice that can help the student to improve his/her writing. This may include referrals for language training and academic writing. The supervisor will provide guidance on technical aspects of writing such as referencing as well as on discipline specific requirements. Detailed correction of drafts and instruction in aspects of language and style are not the responsibility of the supervisor.
6. Will support the student in the production of a research report, dissertation or thesis. Provision should be allowed for adequate, mutually respectful, discussion around recommendations made.
7. Will assist with the construction of a written time schedule which outlines the expected completion dates of successive stages of the work.
8. Will ensure the student has the opportunity to present work at postgraduate/ staff seminars/national/international conferences as appropriate.
9. Will assist with the publication of research articles as appropriate.
10. Will discuss the ownership of research conducted by the student in accordance with the University guidelines and rules on intellectual property, co-authorship and copyright.
11. Will ensure that the research is conducted in accordance with the University's policy on plagiarism.
12. Will ensure that the student is made aware in writing of the inadequacy of progress and/or of any work where the standard is below par. Acceptability will be according to criteria previously supplied to the student.
13. Has a duty to refuse to allow the submission of sub-standard work for examination, regardless of the circumstances. If the student chooses to submit without the consent of the supervisor, then this should be clearly recorded and the appropriate procedures followed.

THE STUDENT:

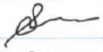
1. Undertakes to work independently under the guidance of the supervisor. This includes reading widely to ensure that the literature pertinent to his/her chosen topic has been identified and consulted.
2. Is obliged to make appointments to see the supervisor and will arrange meeting times well in advance.
3. Will think carefully about how to derive maximum benefit from these contact sessions by planning what he/she wants in these sessions.
4. Should submit written work for discussion with the supervisor well in advance of a scheduled meeting. The kind and frequency of written work should be agreed with the supervisor at the outset of the research.
5. Undertakes to submit written work that is relatively free of basic spelling mistakes, incorrect punctuation and grammatical errors. Responsibility for the accuracy of language, the overall structure and coherence of the final research report, dissertation or thesis rests with the student.
6. Undertakes to heed the advice given by the supervisor and to engage in discussion around suggestions made. Ultimately the student has to take responsibility for the quality and presentation of the work.
7. Should strive, within reasonable bounds, to maintain a focus on his/her research area and to work within the agreed time schedule.
8. Will prepare material for presentations at seminars and conferences.
9. Undertakes to submit papers for publication.
10. Agrees to honour agreements about ownership of the research and in accordance with the University's guidelines and rules in relation to co-authorship, copyright and intellectual property.
11. Will ensure that the work contains no instances of plagiarism and that all citations are properly referenced and that the list of references is accurate, complete and consistent.
12. Agrees to work in accordance with the criteria of acceptability as supplied by the supervisor.
13. Undertakes not to place the supervisor under undue pressure to submit work for examination until the supervisor is satisfied that it has reached an acceptable level of quality.

I confirm that I have read and understood this statement and agree to be guided by its principles

Name of student : Safiyyah Ismail

Student's signature : 

Name of Supervisor : Dr Selvarani Moodley

Supervisor's signature : 

Name of Co-Supervisor : Professor Rod Alence

Co-Supervisor's signature : _____

The broad area of study is: Algorithms for prediction of hearing loss

Provisional submission date is: July 2020

Degree : Master of Arts in eScience

School : Social Science

Faculty : Humanities

Date : 20/07/2020

Specific agreements pertaining to: ownership and joint publication, funding, etc. may be attached and signed.

GRIEVANCE PROCEDURES. It should be acknowledged that during the course of the research, both students and supervisors can feel aggrieved. In this event, matters should be dealt with as swiftly as possible by the parties involved and, if necessary, the appropriate Postgraduate Coordinators and Committees. There is, in addition, a University Grievance Policy to help guide deliberations. It is available on www.wits.ac.za/prospective/postgraduate.

Appendix D

Permission to Use Datasets

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

the centre for DEAF STUDIES 20th Anniversary

HI HOPES

18th March 2020

Dear Ethics Committee

I, Dr Selvarani Moodley, am the database gatekeeper for the HI HOPES Early Intervention Programme longitudinal dataset. Safiyyah Ismail is a Masters Student (in the field of E-Science) that I am supervising.

This serves to confirm that Ms Ismail has been given permission to use the HI HOPES dataset for secondary data analysis for her Master project titled: Detecting Hearing Loss in High Risk Neonates Using a Machine Learning Algorithm.

Please contact me if there are any queries

Regards

Dr S. Moodley

083 234 9826