



# DATA COLLECTION: BEST PRACTICES AND TECHNIQUES OF FOOTPRINT COLLECTION FOR NEWBORN IDENTIFICATION

Enakshmi Ghosh<sup>1</sup>, Tandra Kaviraj<sup>2</sup>

<sup>1</sup> M.Tech, Department of Computer Science and Engineering, Maulana Abul Kalam Azad University of Technology, WB, India

<sup>2</sup> Ph.D Scholar at Nims University, Rajasthan, Assistant Professor at Techno India University, Nursing Superintendent, Techno India DAMA Healthcare and Medical Centre

**Abstract:** Every research is dedicated for the welfare of the mankind and to serve the humanity. Similar to every research that has culminated dreams to reality this research also aspires to serve mankind at its best. As children are destined to be the future citizens of the world, protecting them from premature death or loss from family is a duty to every responsible human being. The domain for this research is in the area of biometric recognition of newborn and infants to preserve their identity. After extensive study on related articles and research materials in this field, a new approach of infant recognition using advanced techniques of computation has been identified. As it is well known that every research to be successful needs appropriate data. But infants biometric, being sensitive data, availability of these data for public research purposes is rare. For the study the researchers gathered 30 samples for the purpose of this study, The paper focusses the data collection techniques, demographics and issues and challenges during the data collection phases. The paper concludes with some discussion on the key factors that can help resolve most of the issues. The researchers hope that this paper can act as a resource of lessons learned for many future researchers.

**Index Terms - Data Collection, Demographics, Child Identification.**

## I. INTRODUCTION

Data is a term used to describe any information that has been predominantly gathered through an observational method. Data collection is the process of acquiring data with the goal of understanding more about the research challenge. Different data kinds correspond to various data collection techniques. However, based on the type of data that is used in the research, it may be tough for researchers to choose the most suitable form of collection method. [1] Data collection is an important stage in research. It can impact the quality of achieving results by reducing the rate of errors which may occur during a research project. Along with a good design for the study, plenty of quality time should be spent in the collection of data to achieve correct results. Insufficient and inaccurate data never guarantees high accuracy of results. [2] Before approaching a data collection

method, the type of data that is required for the study should be determined. There are two types of data for research. They are a) Qualitative data b) Quantitative data

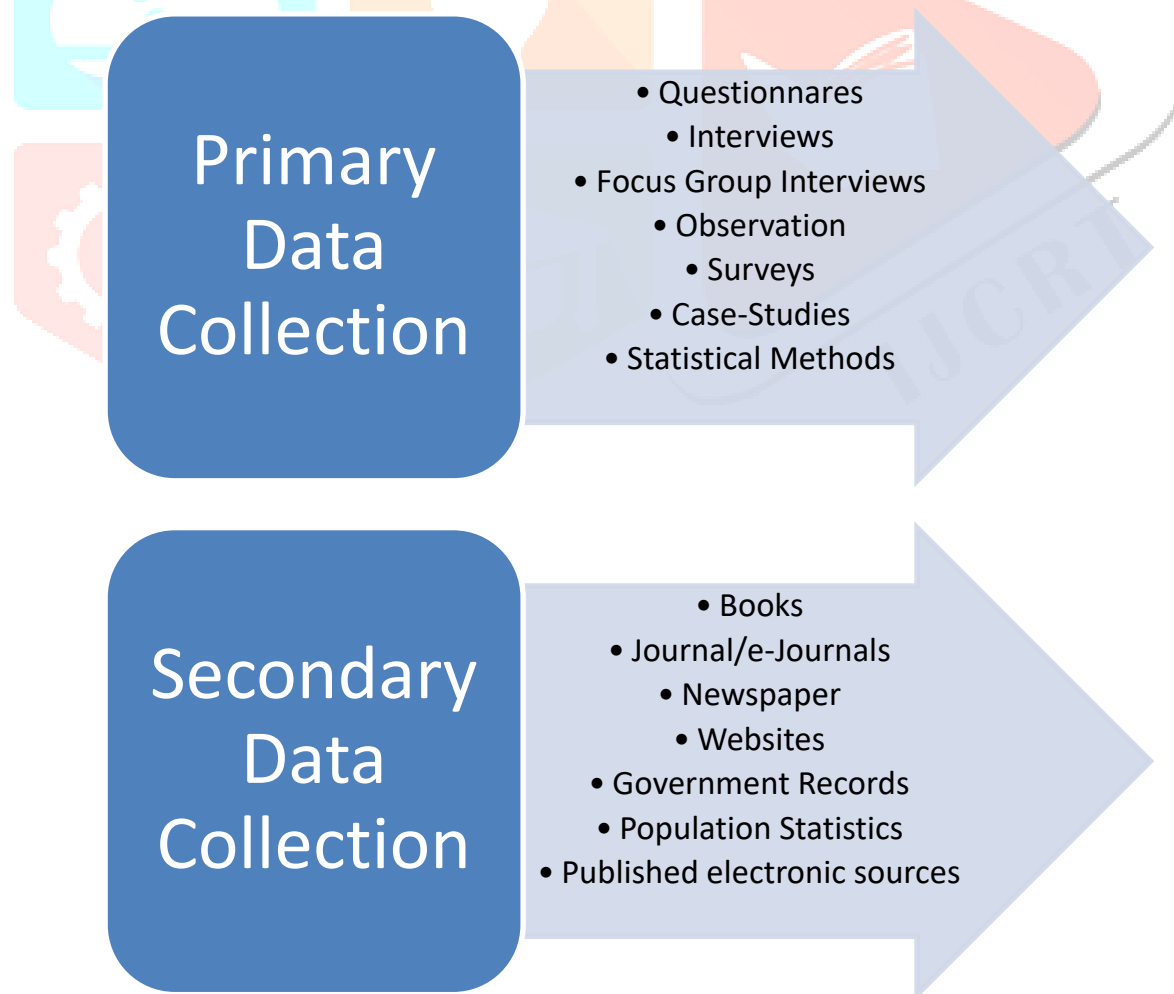
## 1.1 Data Types Used in Research:

Data can be either qualitative or quantitative in nature.

- Qualitative data is typically gathered from a variety of sources but it is difficult to quantify or measure quantitatively. This type of data are used to provide answers to the "how and why" questions in a research project. They mostly consists of information on feelings, perceptions, and emotions ... Researchers obtain these data utilizing a variety of techniques, including audiotapes, sketches, notes, and capturing images. There are two additional categories for qualitative data:(I) ordinal data and (II)nominal.
- Quantitative data are the numerical values of the variables, which represent the quantitative data. Quantitative data is mathematically generated or computed. There are different scales for measuring quantitative data including nominal, ordinal, interval, and ratio scales. They address the “what” question type in a study.(I)Discrete data and (II)continuous data are two categories into which quantitative data can be categorized. Both qualitative research data collecting and quantitative research data collection are crucial to the field of research.[2]

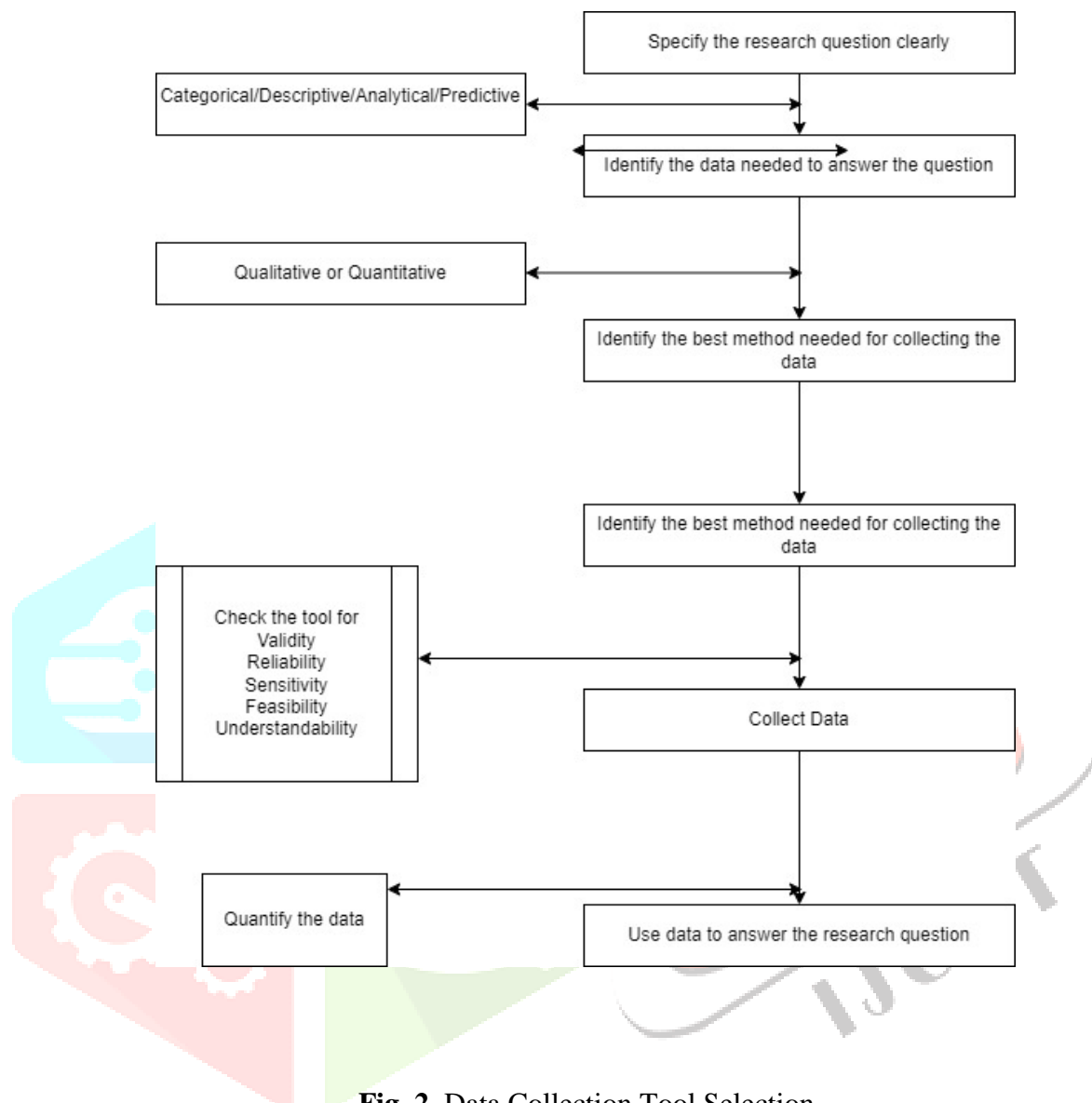
## 1.2 Data Collection Methods And Tools:

A suitable data collection method must guarantee that the collected data have to be accurate and well defined. Generally, data collection methods are divided to two main categories of Primary Data Collection Methods and Secondary Data Collection Methods. Primary data is original information that is obtained directly from a source and is not altered by an outside source. Secondary data is information obtained from published sources, which means it has already been gathered for another cause and may be used for other research reasons. Well known methods of data collection are -Open-ended survey, Quizzes, Interview, Focus-group, Direct Observation[1]. Figure below shows some of data collection methods



**Fig. 1.** Data Collection Methods

Note: Every data collection should take the ethical considerations. Thus, confidentiality of data must be considered [1]. Ethical clearance processes are used to ensure that a project is constructed in accordance with the necessary ethical guidelines. Depending on the type of data and purpose of research, the researchers should select the appropriate data collection strategies. The diagram below illustrates the procedure that every researcher should follow before choosing an appropriate tool for data collection.[3]



**Fig. 2.** Data Collection Tool Selection

This research is based on a case-study approach where data collection strategies for a couple of scenarios has been considered. The case study focuses on the process of data collection, tools used for data collection and concludes with demographic distribution of data. Since the case study was focused on infants below a certain age limits, we have used the techniques of Focus group study for the purpose of data collection for this work. The data collection tool was moderator and responder protocol for data collection. We, begin with understanding the importance of Data Collection for Infant Biometrics.

## II. Literature Survey

In order to understand the importance of Data Collection for Infant Biometric study an extensive study was conducted. It was found that since decades a very interesting questions have daunted the research communities. Can the children in the age group of 0-4 years be authenticated using biometric traits? Biometric data capture pro-grams for NCIP(National Child Identification) or Aadhar is pertained to population in the age of typically 5 years old. In practice, the biometric data are not captured for kids below 5 years in Indian Aadhar program. Even though they are allotted a unique id, the biometric data are actually captured when

they turn 5 years old. They are reenrolled in Aadhar [4]when they turn 15 years old. The primary reason is that biometric traits are fundamental to any human being. But for children the popular biometric traits like iris, face and finger is not stabilized until they turn 15. Due to large variations in the unique characteristics the researches in the area of infant biometrics is quite limited. [5]. But with the rising need of biometric research in the area of infant biometric is quite high. The applications for this research topic are multiple

- Tracking a child for vaccination schedule
- Prevention against newborn baby swapping
- Generation of civilian –id

The above listing is only a subset of areas where infant biometric researches are required. The actual scope covers a wider application arena. But over decades several factors have hindered the advancement of research in this particular area. In the young children in the age group of 0-4, the commonly used biometric traits are not persistent. They are subjected to large variations in a limited span of time. The data collection for research purposes is also hazardous due to the fact that children are generally uncooperative, do not follow instructions thus image cannot be properly captured. The parental consent played a major factor where the parents fear data sensitivity in consideration. Also, a subject available at first session becomes unavailable for subsequent sessions due to lack of importance given by parents for this re-search purpose. Hence most of the research studies focused on adult biometric traits in multiple areas like civilian registry system to mobile device authentication.[6]

Recently with the emerging technology and availability of modern age instruments and techniques some investigations have divulged in the area of footprint as the focal point of study. While uniqueness and persistence have been studied for three most popular biometric modalities—fingerprint, face, iris for adult population there have been no significant studies on footprint as a biometric trait for the area of research in adults.

With the emerging biometric application requiring recognition of infants and toddlers (0-4) the research community as well as healthcare, government agencies are interested in searching the answer to the question—“Can children in the age group of 0-4 years be reliably recognized over time using footprint as their unique biometric traits?”[7]

Our research is primarily focused around this question from the research community. The researchers have reviewed the related studies. In this paper we detail our initiative to study the persistence of footprint for children in the age group of 0-4 years. The uniqueness for our study is that we have included mother’s thumbprint as another key determining factor for generating a perfect match. The key issues, methods and outcome are described in the next section. Based on preliminary results obtained for data collected in our study, we use infant footprint and mother’s fingerprint in the age group of 0-4 years old children that appears to be promising.

Data related to any infant related study is subject to several issues like security, sensitivity, lack of parental consent thus leading to several incomplete data sets for research purposes. Infant data are also very rarely available for public use in re-search arena. Past studies suggested that the researchers mostly gathered data out of their own initiative for the purpose of their study. The table below summarizes the dataset gathered by past researchers for their studies.

**Table 1.** Comparison of Data Set Gathered by Past Researchers and Investigators

Authors	Year	Modality	Subjects	Age	Time Interval
<b>Protichi Basak et al [9]</b>	2017	Fingerprint,Iris,Face	100 children	18 months to 4 years	Two different sessions, four months apart
<b>Sahar Siddiquiet al[25]</b>	2018	Face	Newborns Database (96 -204 newborns)	one hour - few weeks after birth	Two sessions (months apart)
<b>S.Balam et al [7]</b>	2013	Foot	240 images from 40 newborns (images for training=40 and testing= 200)	Two days following births	
<b>Al-Dulaimi et al [3]</b>	2013	Foot	60 images with 40 persons by 8 samples		
<b>Eryun Liu et al[10]</b>	2017	Foot	60 subjects	1 to 9 months	Three seasons

In span of researches the face, fingerprint, and iris modalities of more than 100 infants (ages 18 months to 4) are collected over the course of multiple sessions that took place months apart. The researchers prepared the multimodal biometric data-base of young toddlers as a result of the research In the table below, the face, finger-print, and iris photographs of more than 100 kids from two different sessions are included in the proposed CMB database The research gained some importance in the fact that it provided the first open access multimodal database that offers baseline results for the fusion of face and fingerprints, depending on the application situations.

### III. Data Collection Phases:

Thus, it can be seen that data collection plays a very important role in the newborn biometrics research areas. In this paper a data collection initiative has been organized for the purpose of data gathering of Infant footprints. The sam-ple size was 30 infants in the age group of 0-5 years. The researchers initiated a biometric data collection effort at 117 Lock Gate area at Chetla, Alipore area of Kolkata in collaboration with DNASWS, a non-governmental organization (NGO) working in this area. In the first phase of data collection conducted in 29th January 2023 footprint of 30 infants were collected for left and right foets. These images were accompanied by left and right thumbprints of correspond-ing mother. o study the growth factor modeling for footprint recognition for in-fant we plan to collect data from the same subjects at two more additional time – Phase II(February 2023) and Phase III(March 2023).



### 3.1 Phase I –Data Collection Protocol

Phase I of data collection took place over two days(Jan 27 and Jan 29).Data was collected at a primary school named Derozio school under the supervision of a team of volunteers from DNASWS.

- As a preparatory phase a general awareness campaign was held for the parents. As a result, the parents also willingly cooperated in the data collection process.
- A data collection station was set up manages by a M. Tech student who was also an active member from DNASWS doing project in footprint biometric study. The data station was equipped with Samsung Galaxy A8 Tab, inkpad and paper for thumbprint.
- The child's name, age, gender, address and mother's name along with parents contact number were collected to contact the parents for follow-up data collection during our subsequent phase of data collection.
- Fig shows the footprint images of subjects, their parents and data capture stations inside the primary school. Parents were required to sign a consent form (an initiative of DNASWS committee) giving their consent to provide their child's footprint and their fingerprint images.
- As an incentive containing a packet of powdered milk and a 250 gms of Sabu dana amounting to Rs 64 per subject
- For phase I data collection, each subject was handed a coupon mentioning the date of 2nd visit signed by DNASWS authorized signatory.

**ছাড়পত্র**

সুধী,  
কোড পিঙ্ক প্রজেক্টের জনস্বার্থে আমার হাতের ছাপ ও  
আমার বাচ্চার পায়ের ছাপ আমার পূর্ণ সন্মতিতে  
গ্রহণ করা হয়েছে।  
কুপন ত্রমিক নং.....  
নাম.....

তারিখ..... অভিভাবক স্বাক্ষর.....

**তথ্য অনুসন্ধান ফর্ম**

নাম.....  
অভিভাবক নাম.....  
ছাপ নেবার প্রথম তারিখ.....  
বাচ্চার বয়স.....  
ছাপ নেবার দ্বিতীয় তারিখ.....  
বাচ্চার বয়স.....  
স্থান.....

কুপন নং-  
নাম.....

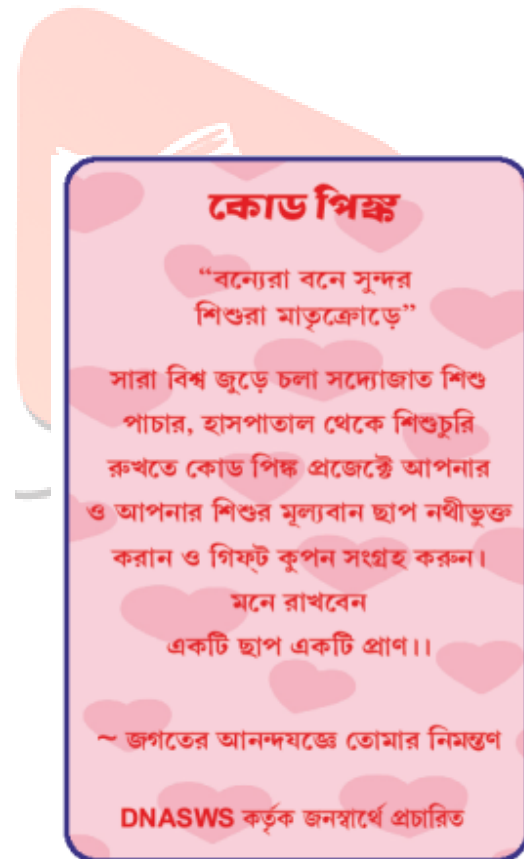
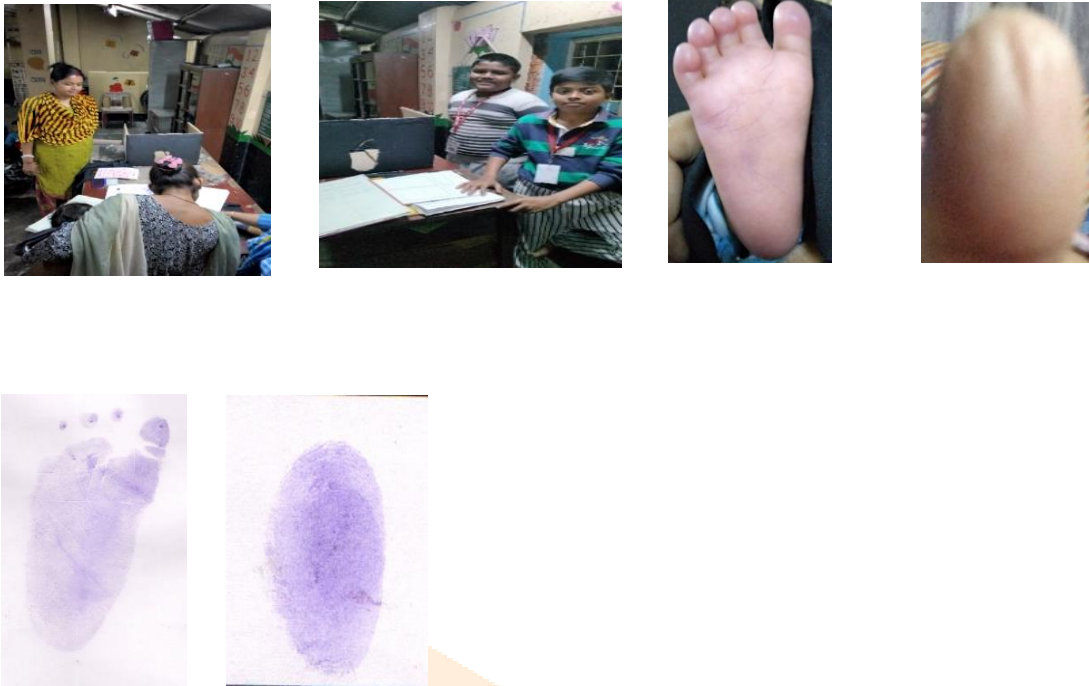


Fig. 3. Pamphlet Samples for Project Awareness (b) Parental Consent for Project Awareness



**Fig. 4.** Data Collection effort Phase I

Figure Description :(a) Parents signing no objection consent form to provide their child's biometric data (b) data collection stations in the school manned by DNASWS volunteers (c) Sample footprint (infants) and fingerprint images from the dataset that was collected (Mobile Capture) (d) Sample footprint (infants) and fingerprint (mothers) images from the dataset that was collected (Offline Capture using ink and paper)

### 3.2 Phase II –Data Collection Protocol

The phase II of data collection took place on 5th Feb 2023. It was again scheduled on 10th February, 2023 to ensure 100% attendance of all subjects.

- To retain the attraction of subjects DNASWS organized a small talent hunt at school premises (SIT-N-DRAW and GO-AS –U –LIKE). Thus the subjects were interested to attend the event alongside participate in the data collection event. Picture below shows some event photographs.
- The data collection at Phase II were only footprint (left and right) images of the child captured on mobile.
- Similar to the first phase in the second phase also provisions were kept at the collection stations for offline data capture using ink and paper for both mother and child. This was done i) as a proof of official record for DNASWS and ii) for future research requirement in this area (if any)
- The data collection of phase II was done using (a) SAMSUNG GALAXY TAB 8 (b) a cardboard box covered with black chart paper. A footprint outline was cut. The infants were made to sit inside the box. They let their legs out of the hole in the box (left first then right). This was done to ensure a black background for images.
- The phase II of data - collection ended with 80% attendance. Prizes were distributed for events on the same day.
- For attending phase II of data collection each subject was given incentives double the first time to retain attraction (2 milk packets-1 mother and other for child), 1 packet of cornflakes for the child and a packet of Sabu dana for mother. The cost of incentive was Rs. 84 per subject)

The incentives were handed over to each subject mentioning the date for final visit. Each subject was also informed about the attractive mega gift hamper waiting for them for final day. This was done to ensure 100% attendance. The aim of the collectors was 100% complete data set in all 3 sessions. In this phase also DNASWS authority and volunteers actively volunteered the data collection process.



**Fig. 5.** Data Collection effort Phase II

*Description of Data Collection Phase II at Derozio School Chetla Lock gate area, Kolkata, West Bengal (a) Sample footprint (infants) that was collected (Mobile Capture) (b) Sample footprint (infants) and fingerprint (mothers) images from the dataset that was collected (Offline Capture using ink and paper) (c) event organized for entertainments at another room in the school (d) parents actively involved in data collection*

### 3.3 Phase III -Final Data Collection Protocol

The final day of data collection was conducted on 5th March, 2023, a month after the enrollment phase.

- The data collection of final phases was done using
  - (a) SAMSUNG GALAXY TAB8
  - (b) a cardboard box covered with black chart paper.
- The mobile capture and offline inked footprints were verified with the previous samples to ensure better accuracy.
- The verified subjects were asked to assemble at the gift distribution arena.
- The gifts were distributed to 30 participating subjects under the initiative of DNASWS representatives and volunteers.
- The gift hamper consisted of a double bed mosquito-net for newborn and the mother. Along with it 250 gms. of Sabu dana were gifted to each subject. The distribution event lasted for an hour. Each participant was asked to sign a received register in agreement of the gift that they received.
- The subject participation was 100% to ensure completeness of the acquired data.





**Fig. 6.** Data Collection effort Phase III

Figure Description-- Data Collection effort Phase III at Derozio School Chetla Lock gate area, Kolkata, West Bengal (a) Sample footprint (infants) that was collected (Mobile Capture) (b) Sample fingerprint (mothers) images from the dataset that was collected (Offline Capture using ink and paper) (c) Mega Gift Hamper handed over to subjects in their final day visits.

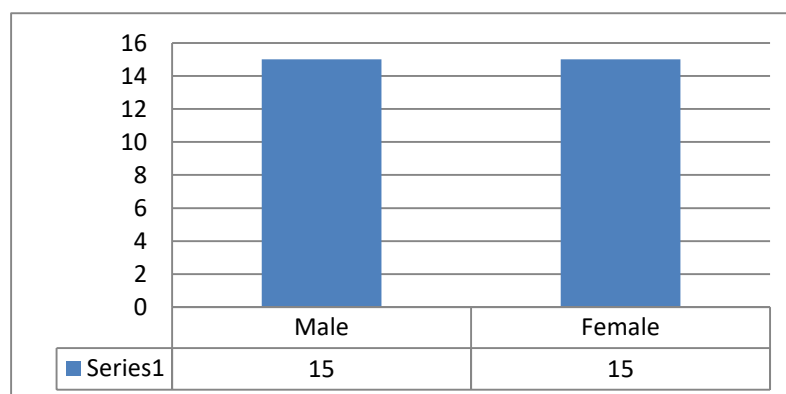
### 3.4 Demographic Distribution of Data

**Table 2.** Out of 30 subjects who provide their biometric data, there are 15 (50%) male and 15 (50%) female across all three sessions of data collection.

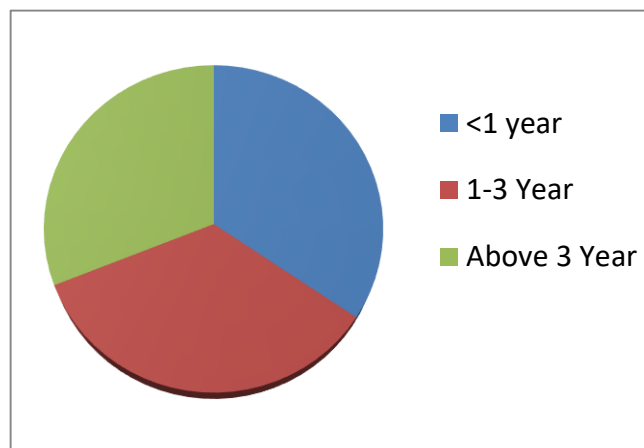
Gender	Session 1	Session 2	Session 3
Male	15	15	14
Female	15	15	15
Total	30	30	29

**Table 3.** Out of 30 subjects who provide their biometric data, there are 15 (50%) male and 15 (50%) female across all three sessions of data collection.

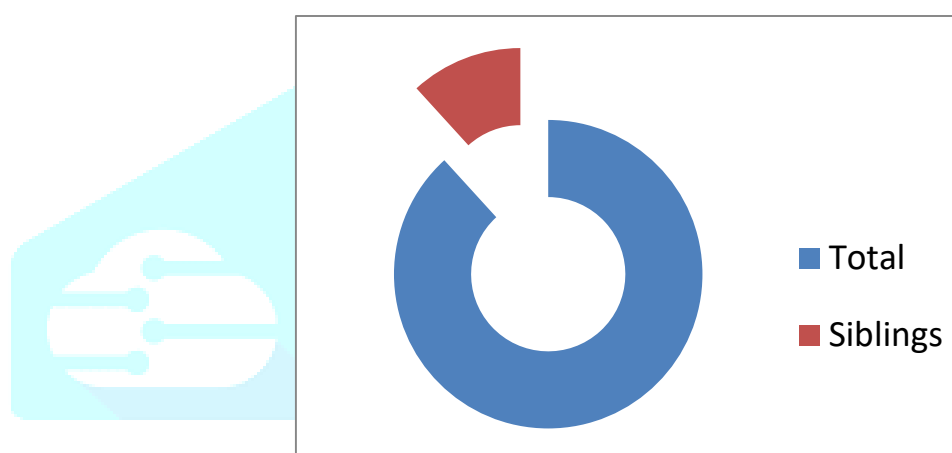
Age	<1 year	1-3 year	4 -5 year
Male	6	5	4
Female	4	5	6
Total	10	10	10



(a)



(b)



(c)

**Fig. 7. Demographic Distribution of Data**

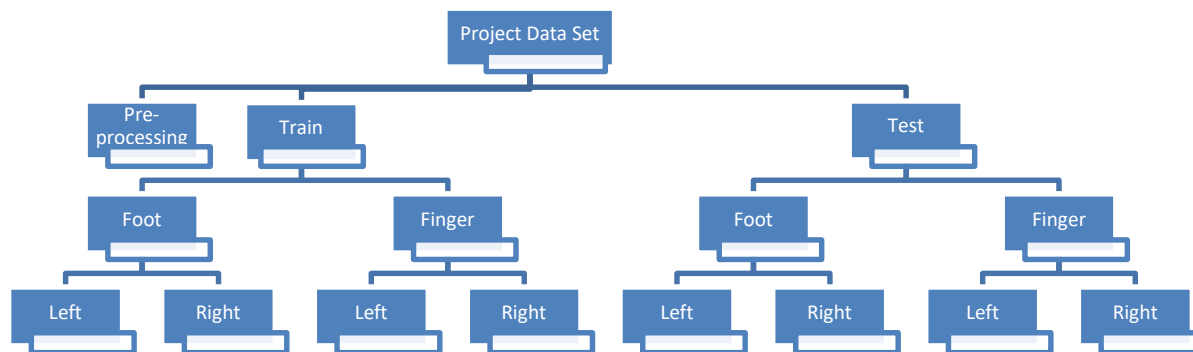
**Figure Description :**(a) Age distribution (b) Gender (c)Siblings Ratio of 30 subjects who appeared for Data Collection.

**Note: Siblings Ratio: Out of the 30 subjects, 2 pairs of siblings are found.**

### 3.5 Dataset Information

The nomenclature followed in the collected samples are as is as follows.

modality \_ visitNo \_ sampleID : This nomenclature is followed for all the three sessions of Data Collection .sampleID is taken at the time of subject registration, modality is either fingerprint or footprint, visitNo is the session identification number e.g. RTF\_V1\_S1.



**Fig. 8.** Folder Structure of Data Set

**Data Availability.** All data is available from the authors.

### Key Points

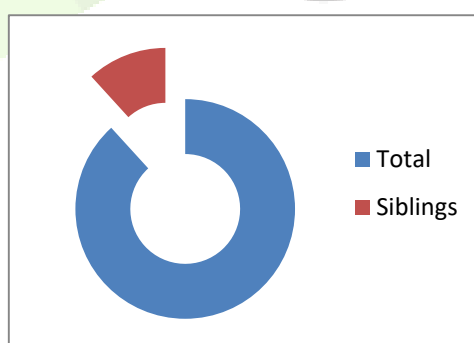
- . The mega gift hamper consisted of a mosquito net worth Rs.200 and 250gms sabudana worth Rs.64. Thus each sample were given a gift worth  $(Rs200 +Rs. 64)=Rs264$  funded by DNASWS.
- The data collection was a major success where 99% of the subjects appeared for final footprint data acquisition thus ensuring complete data set of 30 samples.
- Only 1 subject was absent due to health issues.
- The gift hampers had a great impetus among all subjects and few of them were willing to refer some more subjects to participate.
- At the end it must be noted that further data collection had to be put on hold for some time due to sudden spread of Adena virus leading to unavailability of infants.

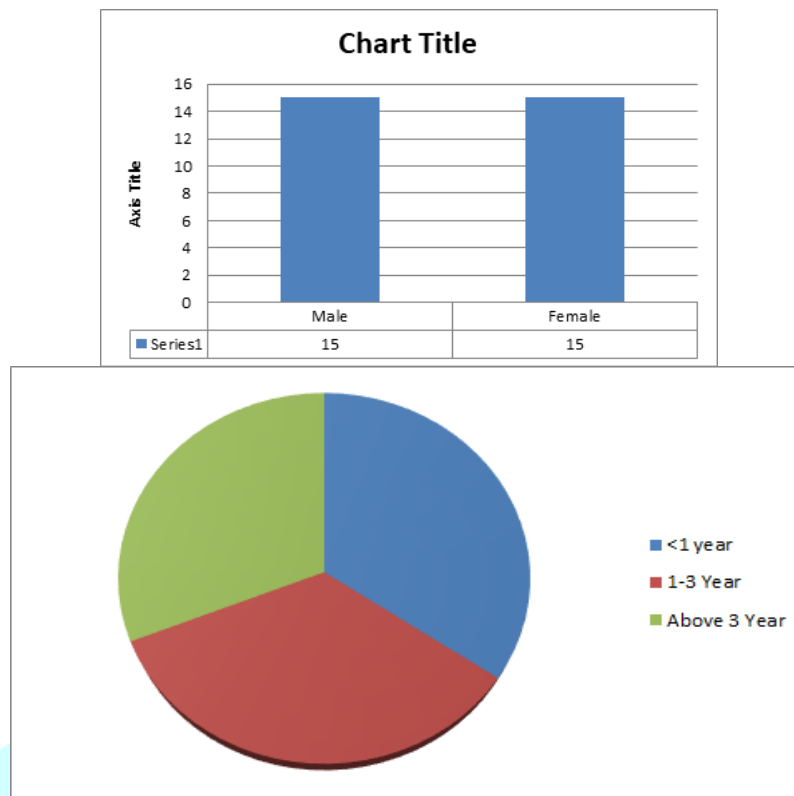
### 3.6 Demographic distribution

**Gender.** : Out of 30 subjects who provide their biometric data, there are 15 (50%) boys and 15 (50%) girls.

**Age:** The age distribution of age is shown in chart below. There are 10 subjects in the age group lesser than 12 months ,10 subjects in the age group of 1-3 years and 10 in the age brackets of 4-5 years.

**Siblings.** : Out of the 30 subjects, 2 pairs of siblings are found.





**Fig. 9.** (a)Siblings Ratio (b)Gender and (c) Age distribution of 30 subjects who appeared for Data Collection

## IV. Challenges and Observation in Data Collection

### 4.1 Issues and Challenges in Infant Biometric Data Gathering

According to the Literature survey proposed by Tempestt Neal et al[9] there are several barriers to biometric identification for infant and toddlers that re-mains a challenging issue even to this date. This paper also would like to bring certain points to highlight the roadblocks in Data Collection for this work and the avenues to surpass them.As this is well known that the Child identification programs (CIPs), like National Child Identification Program (NCIP) ,Indradhanush<sup>1</sup> were designed to track Child Vaccination schedules, Tracking Child Vaccination Schedules and Identifying Missing Children only to name a few. A cross-sectional survey to identify General Views of Infant and Toddler Identification revealed that there are several issues concerns with biometric collection that has been graphically represented in the study conducted by Tempestt Neal et al[9].Of the survey 7% population was concerned for Data Security issues,10% distrust in government ,10% distrust in biometric system and 17% feared the long term impact on infants.Thus the researchers in the survey conducted by Tempestt Neal et al[9] suggested there are certain remedies that can help remove the issues in the long term like

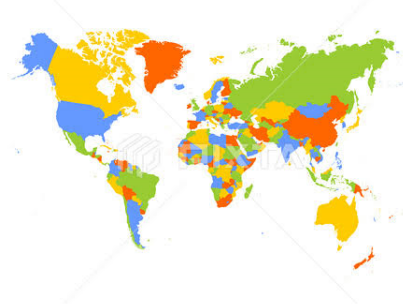
- Improved visibility on NCIP programs from the birth of the child among new parents
- Clarity on privacy measures among parents
- Providing operation control on parents
- The temporary storage of data

According to their survey conducted there are several barriers to bio metric identification for infant and toddlers that remains a challenging issue even to this date. This paper also would like to bring certain points to highlight the roadblocks in biometric data collection for this work and the avenues to surpass them. There are certain remedies that can help remove the issues like

<sup>1</sup> <http://www.nhp.gov.in/mission-indradhanush>

- 1.Improved visibility on NCIP programs from the birth of the child among new parents
- 2.Clarity on privacy measures among parents
- 3.Providing operation control on parents

Assuring temporary storage of data



Issues	Percentage
Physical Changes Over Time	3%
Distrust in Biometric Systems	10%
Long Term Consequences to Infant	17%
Data Security	7%
No Consent from Child	23%
Distrust in Government	10%
Privacy	3%
Parent Consent	14%
Unethical	3%
Enrolment Problems	10%

**Fig. 10.** Graph showing general issues and concerns on Biometric Capture for Infants[9]

## V. Challenges in Data Collection

While preparing for data collection the researchers kept in mind some of the key issues faced by past data collectors. Accordingly, we planned certain activities or precautionary measure to eliminate or handle them.

In this section, we discuss some of techniques adopted by current researchers.

### *General Awareness Building*

Before data collection a pamphlet was designed for building general awareness on the issues and benefits of attending the session. Pamphlet campaigning in the area of the data collection generated positive response.

### *Parents Interactive Session*

Before data collection started an interactive session was conducted with a facility of one to one interaction. general and specific queries from parents and guardians were conducted for 30 minutes each day during Phase I.

**Parents Involvement.** Parents are allowed to actively involve in the data collection initiative. Since the process was made easy it was possible for parents to support their kids during data capture. Also they got very enthusiastic when given a chance to participate. This reduces the general disagreement/disapproval or opposition from parents to a great extent. They in turn took an active part to convince more parents to participate in collection.

**Incentive:** Incentives are divided into 3 set-one for each session with from parents mega gift for attending the final session. This helped retaining the interest of the subjects, thus ensuring 100% attendance resulting in complete data set.

**Event :** A small event was organized to entertain the waiting queue. Thus, parents and their children could participate as well as not bored of waiting in the collection procedure.

But in spite of above initiative certain challenges still remained. They are

- **Dirty Foot:** We observed that foot of few subjects were dirty. In such cases we cleaned the subjects foot with Dettol and sanitizer. For one subject it was painted with red paint. We requested the accompanying parents to wash their child's foot with soap and water before Phase III of final data capture session.
- **Dry Finger:** Due to the fact that data collection was held in winter due to cold and dry environment many newborn mother's had skin peeling and dry skin issues. Dry fingers and peeled skins hide many of skin ridges leading to poor quality of images or thumb impressions. We washed their hands with wet wipes to enhance picture quality. [8]
- **Manual vs Autocapture:[8]** Due to unavailability of Fingerprint Reader during data gathering, we used inkpads for gathering thumb impression of the subjects. In our experience automatic capture was more convenient and efficient for biometric gathering purpose. We expect future researchers to use fingerprint-reader for auto-capturing of fingerprint for more accurate researches.



- Fear among children: Some subjects though very few were quite uncomfortable because they were being brought to a different environment. They were calmed down by their parents and became interested when they saw other children who comfortable. They pacified and calmed down.
- Age Calculation: Since most of the subjects were from slums few mothers were unable to give exact age in terms of days and months. Since data was captured in a span of few weeks most parents were unable to provide exact date, days for the children. For them the collectors asked them to bring the birth certificate of the children and helped them to calculate the same in the proper manner.

## VI. Conclusion and Future Scope

To conclude, the paper focusses the need of data collection for research studies. It describes the scenarios how infant biometric data are collected in 3 sessions, the ways to gather complete data set and how to deal with issues that are faced by past data collection efforts in this area. As a future scope for this work, we plan for processing the data set and use learning models to test for prediction accuracy. The researchers also plans to expose the data set as an open access for future research studies thus helping other researchers to start from the implementation phase without having to collect data samples from the scratch.

## VII. Acknowledgement

The author would like to thank the residents of LockGate Area,Chetla,Alipore,Kolkata for their warm and active support in the data collection process and DNASWS who provided their volunteers and logistics support for this study. They would like to thank the owners and administrators of Derozio School They would also thank Nursing Super & Operations Manager,Techno India Dama Hospital for their encouragement and support for this research

## REFERENCES

- [1] Taherdoost, Hamed. "Data Collection Methods and Tools for Research; A Step-by-Step Guide to Choose Data Collection Technique for Academic and Business Research Projects." *International Journal of Academic Research in Management (IJARM)* 10.1 (2021): 10-38.
- [2] Kabir, Syed Muhammad Sajjad. "In book: Basic Guidelines for Research: An Introductory Approach for All Disciplines (pp. 201-275)." (2016).
- [3] Wade, Derick T. "Assessment, measurement and data collection tools." *Clinical rehabilitation* 18.3 (2004): 233-237.
- [4] Unique Identification Authority of India. <http://uidai.gov.in/>.
- [5] Kotzerke, Johannes, et al. "Steps to solving the infant biometric problem with ridge-based biometrics." *IET Biometrics* 7.6 (2018): 567-572.
- [6] Jain, Anil K., et al. "Biometrics for child vaccination and welfare: Persistence of fingerprint recognition for infants and toddlers." *arXiv preprint arXiv:1504.04651* (2015).
- [7] D. Weingaertner, O. R. P. Bellon, L. Silva, and M. N. Cat. Newborn's biometric identification: Can it be done? In *VISAPP* (1), pages 200–205, 2008
- [8] Basak, Protichi, Saurabh De, and Mayank Vatsa. "Biometric recognition for kids." (2017).
- [9] Neal, Tempestt, and Ashokkumar Patel. "A Brief Literature Review and Survey of Adult Perceptions on Biometric Recognition for Infants and Toddlers." *2020 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2020.