Extending Shabda Pariksha using Machine Learning

Mentor code: BJS1_M120   **Mentor**: Dr. NATARAJ K S
Intern: XXXXXX

AUGUST 2022

Performed under the
IKS INTERNSHIP PROGRAM

TABLE OF CONTENTS:

ABSTRACT

Traditional Ayurveda uses eight types of clinical examination (called Ashtavidha Pariksha) for i) disease diagnosis and ii) understanding the composition/condition of any individual. One amongst the eight is Shabda Pariksha - which is an examination based on voice. Based on voice quality, the three doshas (Vata, Pitta, and Kapha) can be predicted according to Ayurveda. Nadi Pariksha (which is also another type of clinical examination amongst the eight types, which is based on the pulse) is also used to predict different psychological bents (likings and dislikings), passions, soft skills, and other personal traits of the individuals, apart from disease diagnosis and prognosis. However, similar predictions are not seen in the literature in the case of Shabda Pariksha. The current project examines the relationship between voice quality and the three doshas, and it extends Shabda Pariksha to understand the correlation of voice with different personality traits, which include likes and dislikes in terms of i) colors ii) subjects iii) types of music iv) different climatic conditions v) tastes, vi) types of movies and vii) hobbies. Voice samples are collected from people belonging to different backgrounds along with answers to a questionnaire for quantifying the three doshas and all the personal attributes mentioned above. A database providing the voice samples, the corresponding amount of three doshas of the person, and the corresponding personality traits is not currently available. Such a database will be extremely useful in systematic study of several aspects of Shabda Pariksha. Thus, in the first month of the internship, we have collected data from 100 subjects, including students, research scholars, and faculty. The data was collected in an anechoic chamber in the studio setup of the Indian Institute of Information Technology Dharwad. The setup has a good quality microphone and isolated chamber to obtain good quality audio recordings. In the next phase of the internship, the correlation between acoustic features (such as pitch, intensity, spectral characteristics) and recorded personal traits will be studied to derive some conclusions. We will also be collecting additional data, and training machine learning models to obtain an automatic predictor which can finally predict the amount of doshas and attributes based on any given new voice sample.

EXECUTIVE SUMMARY

More than 3000 years ago in India, the world's oldest holistic healing system was developed. Commonly known as Ayurveda, this system is based on the belief that health and wellness depend on a delicate balance between the mind, body and spirit. According to these beliefs, there are three main doshas in an individual: vatta, pitta and kapha. These doshas are based on a person's physical, emotional, mental and behavioural characteristics of the person.

Vata consists mostly of the two elements air and space (also known as ether) and is generally described as cold, light, dry, rough, flowing, and spacious. Autumn represents vata for its cool, crisp days.

Kapha (pronounced "kuffa") is based on earth and water. It can be described as steady, stable, heavy, slow, cold, and soft. Spring is known as kapha season, as many parts of the world slowly exit hibernation.

Known for being associated with a tenacious personality, the pitta dosha is based on fire and water. It's commonly described as hot, light, sharp, oily, liquid, and mobile. Summer is known as pitta season for its sunny, hot days.

Under the IKS internship carried out over the past two months, we aimed to predict a person's vatta, pitta and kapha doshas using various other traits they possess. The point of the study was to create a machine learning or a deep learning system that would take in a person's audio parameters as input and predict the vatta, pitta and kapha scores of the given person. Usually, only of the three would have a marginally significant score as compared to the other two. As a result, we know that the individual in question is more of a vata person, or a pitta person or a kapha person.

Carrying out this study involved five interns to work with our mentor. The first thing we did, before beginning the study, was to look into all the existing literature and learn the actual traits that get affected by these doshas described in ayurvedic science. We read several books such as "A glimpse of Ayurveda - The forgotten history and principles of Indian traditional medicine" by Jaiswal YS and "7th Edition, Shabda Pariksha" by "Chaukhamba Publications". Reading these books and watching many related videos gave us a good understanding of these doshas. We made several notes on Ashtavidha Pariksha as well.

We then began creating questionnaires for the purpose of data collection. For this, we made sure to incorporate all essential features into the questions. This includes but is not limited to weight gaining capability (less for vata, medium for pitta and more for kapha), walking and talking speed (fast for vata, medium for pita and less for kapha), type of weather enjoyed, regularity of appetite, skin texture, size of eyes, how easily restless an individual gets, how much one worries, what kind of food they like, what kind of music they prefer to listen to, which type of movies they enjoy, what hobbies they have and so on. There was a significant amount of thought put into building these questionnaires. It was a two part questionnaire along with an observation sheet.

A duration of five to six weeks was then spent collecting the data. We created a setup in our college for the five interns and our mentor to be able to ask various subjects the questions we prepared. They were first made to sign a consent form. We then recorded their audio, while they enunciated the 26 different vowel and consonant sounds in hindi. They then proceeded to answer the two part questionnaire. Collecting data from one subject took approximately twenty to twenty-five minutes. Each subject was then awarded with an incentive as promised, a chocolate or a notebook depending on the availability and choice. Data was successfully collected from more than 200 subjects over this course of time.

The next imperial step in the process of the study was to clean the data whilst working on the model. The intention of the model was to be able to take in audio traits of a person, and predict their vata, pitta and kapha scores. To be able to feed audio traits, we hadn to extract these traits from the audio files to begin with. TextGrid files were created for each subject's audio file. These files were then optimised and corrected by the interns manually, using certain audio processing softwares.

We then had to clean the data obtained from the subjects in the form of questionnaires. While cleaning it, we found several null values and inconsistencies. To fix that, several subjects were called for the second time, and were patiently asked the questions once again. To guarantee the sanctity of the data for the sake of accurate results in the study, we all gave our best efforts to get the cleanest possible data.

Once our dataset was cleaned, sorted and organised in well maintained csv files, we used these questionnaire answers to generate vata, pitta and kapha scores for each of the subjects.

Each question in the questionnaire had a specific answer that fed to vata, pitta or kapha. If the answer given by the subject coincided with the vata option in the question, the vata score of said subject incremented. Likewise, if the answer coincided with pitta or kapha, the pitta or kapha scores respectively incremented. At the end, more often than not, only one out of the three scores

had a significant value as compared to the other two. Thus, it could conclusively be said that the subject in question has the highest vata dosha (or pitta dosha or kapha dosha, likewise).

Now that the audio traits and vata, pitta and kapha scores had been generated, these were fed into the machine learning model. Various regression models were tried and tested for most accuracy. The model finally extracted a relationship between the audio parameters (which were the input features) and the vata, pitta and kapha scores (which were the output features) which aided the accuracy of the predictions.

The objective of the study had been achieved. We will now look into the different steps conducted during the study in detail through the rest of this report.

INTRODUCTION

Ayurveda is one of the oldest systems of medicine practiced in India. It is based on the assumption that the universe is made of five elements (known as Pancha Mahabootha in Ayurveda): Aakash (Space or ether), Agni (Fire), Vayu (Air), Jala (Water), and Prithvi (Earth). These five elements are believed to form three human body types: Vata dosha, Pitta dosha, and Kapha dosha. The physiological functions of the body, mental, and emotional health of the individuals are controlled by the amount of each of these doshas. An imbalance in these doshas results in various diseases in the body [1]. Ayurveda provides several clinical examinations (referred to as Pariksha) for diagnosing diseases: Dwividha Pariksha, Trividha Pariksha, Chaturvidha Pariksha, Panchvidha Pariksha, Asthavidha Pariksha, and Dashavidha Pariksha. The Asthavidha Pariksha involves eight examinations (Nadi (Pulse), Mootra (Urine), Mala (Faeces), Jihwa (Tongue), Shabda (Voice/Sound), Sparsha (Touch), Drika (Eyes) and Akriti (Posture)) is considered as an effective tool [2], [3]. The Shabda Pariksha involves an examination of the quality and nature of the voice to identify different doshas [4]. The three doshas are known to affect a person's physiological, mental, and emotional health. [5], [6]. Thus, it can be derived that the mental and emotional characteristics of the person can be derived from Shabda Pariksha. The Nadi parisksha is reported to help in understanding the psychological states of the person [7]. However, Shabda pariskha is not used in understanding the psychological bents (likings and dislikings), passions, soft skills, and other personality traits of the individuals, apart from disease diagnosis and prognosis.

The goals of this project will be i) to assert experimentally if the voice quality is related to three doshas as described in Shabda Pariksha of Ayurveda, ii) to assert experimentally if an extension of the scope of Shabda Pariksha can be done towards predicting different personal traits based on voice samples, ii) to build a machine learning model to map the voice samples to the personality traits with more accuracy. The outcome will be i) a machine learning model which predicts different personality traits based on the voice samples of any person, ii) a database consisting of all voice samples, the corresponding amount of Tridoshas, and the corresponding attributes of subjects, without disclosing personal details, and iii) a research paper to be presented in reputed conferences.

The current project examines the relationship between voice quality and the three doshas, and it extends Shabda Pariksha to understand the correlation of voice with different personality traits, which include likes and dislikes in terms of i) colors ii) subjects iii) types of music iv) different climatic conditions v) tastes, vi) types of movies and vii) hobbies. Voice samples are

collected from people belonging to different backgrounds along with answers to a questionnaire for quantifying the three doshas and all the personal attributes mentioned above.

A database providing the voice samples, the corresponding amount of three doshas of the person, and the corresponding personality traits is not currently available. Such a database will be extremely useful in systematic study of several aspects of Shabda Pariksha. Thus, in this project we have put significant effort to collect such data with good quality. The project involves data collection, examination of the relation between voice characteristics, three doshas, and different personality traits, and machine learning modelling. The data collection is done at the Indian Institute of Infomration Technology Dharwad. The data is collected at a studio setup of the institute as it helps in collecting good quality audio recordings. The institute has students, research scholars, faculty, and teaching and non-teaching staff making convenient to collect data.

BACKGROUND

Details of the Indian Knowledge System component

In this section, we will briefly describe the physiological and psychological characteristics of three doshas (Vata dosha, Pitta dosha, and Kapha dosha). Further, we will describe details of the Shabda Pariksha.

The physiological and psychological characteristics of the three doshas are described below [5], [8].

Vata dosha

Physiological characteristics: Thin body, dry skin, fine hair, irregular appetite, irregular bowel movements, frequent gas formation, sensitive to cold, difficulty in sleeping.

Psychological characteristics: Energetic with quick action, overstimulated, quick learning, stressful/anxious in difficult situations, can forgive quickly, highly creative.

Pitta dosha

Physiological characteristics: Medium build and muscular, healthy skin and hair, always hungry, frequent bowel movements, higher metabolism, like cool foods, does not like heat.

Psychological characteristics: Aggressive with leadership qualities, impulsive, goal oriented, purposeful learning, strong emotions, and dynamic response to every situation.

Kapha dosha

Physiological characteristics: Strong build, less appetite, regular bowel movements, good sleep, low metabolism, sensitive to cold and damp weather,

Psychological characteristics: Calm and slow to react, easy personality and lovable, slow learning, less motivated, routine life, does not get angry easily, and depressed sometimes.

The basic details of **Shabda Pariksha** [8]

Shabda Pariksha involves examining the voice type and quality. The heavy, sweet, and deep voice are associated with the Kapha dosha. Broken and cracky voices are associated with the Pitta dosha. The rough, hoarse, and disliked voice is associated with the Vata dosha. Shabda Pariksha involves listening to sounds produced by different organs such as lungs, stomach, heart, and other body sounds such as cough and bowel sounds. The abnormal health condition can be diagnosed by classifying voice as normal (the sound similar to swan, the sound similar to demoiselle crane, the sound of felly of wheel, the sound of kettle drum, the sound of house sparrow, the crow sound, the sound of pigeon, the sound of jahrjahra drum) or abnormal voice

(the sound of sheep or goat, the feeble sound, the unclear sound, the choaked type sound, the faint sound, and stammering sound)

Earlier related works

In literature, several studies have been carried out correlate the Tridoshas to psychological characteristiscs [5], [6]. The earlier literature on Shabda Pariksha has been focused on the categorization of voice to various disease conditions [8], [9], [11]. However, there are no studies to correlate voice qualities to different personality traits. Also, no studies have extracted voice features (such as pitch, intensity, and spectral parameters) to quantify the relation with Tridoshas and personality traits.

A database providing the voice samples, the corresponding amount of three doshas of the person, and the corresponding personality traits is not currently available. Thus, a significant effort is put in collecting a good quality dataset. The current project examines the relationship between voice quality and the three doshas, and it extends Shabda Pariksha to understand the correlation of voice features (pitch, intensity, and spectral parameters) with different personality traits which includes likes and dislikes in terms of i) colors ii) subjects iii) types of music iv) different climatic conditions v) tastes, vi) types of movies and vii) hobbies.

DESCRIPTION OF THE PROJECT

The project is carried out at the Indian Institute of Information Technology Dharwad. The data is collected at a studio setup of the institute. The studio setup has an anechoic chamber to avoid echoes in the recorded sound due to sound wave reflections. Also, the room provides very good isolation from the surroundings, helping in collecting good quality audio recordings. The institute has students, research scholars, faculty, and teaching and non-teaching staff, making it convenient to collect a large amounts of data as required for correlation study and machine learning.

METHODS

The project involves data collection, and examination of the relation between voice characteristics, three doshas, and different personality traits, and machine learning modelling. The students, research scholars, faculty, and teaching and non-teaching staff at the Indian Institute of Information Technology Dharwad were requested to be subjects for data collection. The details of the data being collected are explained to each of the subjects, and their consent is obtained to collect their data. The form used for taking the consent of the subject is provided in Appendix I.

The data is collected at a studio setup of the institute, shown in Figure 1, as it helps in collecting good quality audio recordings. The setup uses the Sennheiser EW 112P G4-A wireless microphone to record the audio at a sampling frequency of 44.1 kHz. Two types of audio are recorded from each of the subjects. The first recording involved recording two-minute audio of the subject speaking about his/her memories of school days and examinations. This audio will be used to obtain the average pitch, intensity, and other prosody-related voice features. The second recording involved the subject uttering 36 phonemes in the Hindi language. These phonemes occur in most of Indian languages, and these recordings will help us to quantify differences in

pronunciation of these basic phonemes. The list of the phonemes uttered by the subjects is provided in Appendix II.

To carry out the correlation between the voice features, Tridoshas, and personality traits, a questionnaire is designed based on the earlier literature to quantify the amount of Tridoshas and personality traits. The questionnaire for Tridoshas involved questions about different weathers, food items, sleeping habits, appetite, skin and hair types, cold and cough, headaches, stomach problems, joint problems, back pain, teeth problems, waking time, learning speed, memory skill. The questions about their mental status, aggressiveness about completing a task, and their liking of different subjects are also included to quantify Tridoshas. The questionnaire also involved several questions for acquiring information about their personality traits. These are the questions about likes and dislikes in terms of i) colors, ii) subjects, iii) types of music, iv) different climatic conditions, v) tastes, vi) types of movies, and vii) hobbies. The list of the questionnaire is provided in Appendix III. The data collection procedure involves explaining the project and taking the consent, recording two audio tasks, and recording the answers to questionnaires. The entire process of collecting the data for one subject requires 50 to 60 minutes.

The correlation between the voice features (such as pitch, intensity, spectral parameters), amount of Tridoshas, and personality traits will be studied in the next phase. For this study, the voice features will be extracted from the audio recordings, and responses to the questionnaire will be converted to values representing different Tridoshas and personality traits. Machine learning models will be trained to predict the amount of Tridoshas and personality traits using a part of the collected data. The remaining part of the collected data will be used to tune and test the machine learning model.

The procedure for the VPK scores extracted

A person's vata, pitta and kapha scores were to be predicted using the subjects' personality traits we received from the questionnaires we conducted. The questionnaires were created after a great deal of research on the vata, pitta and kapha doshas along with their effects on a person. The questionnaires captures qualities such as weight gaining capability (less for vata, medium for pitta and more for kapha), walking and talking speed (fast for vata, medium for pita and less for kapha), type of weather enjoyed, regularity of appetite, skin texture, size of eyes, how easily restless an individual gets, how much one worries, what kind of food they like, what kind of music they prefer to listen to, which type of movies they enjoy, what hobbies they have and so on. Each question has a set of options. Each option carried a score for vata or pitta or kapha.

Once the dataset containing the questionnaire answers was cleaned and fed to the script we created, in the form of a csv file each of the questions' answers were iterated through.

A separate csv file was created with each question's reference letter or index along with its answer that contributed to vata, pitta or kapha in the following tabular format.

	Questionnaire	V	P	K
0	D	3	2	1
1	E	3	2	1
2	F	3	1	3
3	G	1	3	3
4	H	100	100	1
5	I	1	3	2
6	J	1	2	3
7	M	3	100	100
8	L	100	100	3

In the above table, it is clear that the question with reference letter 'D' has three options: 1,2 and 3. If the subject answered with 1, then their kapha score gets incremented. Likewise, if they answered with 2 or 3, their pitta and vata scores respectively get incremented.

There were several questions which had answers on a scale of 1 to 5. The answers were modified and brought down to a scale of 3 using the same script. Now that all the answers have been normalized, they were iterated through.

The question's answers were iterated through and each question's index was checked for in the latter csv file. The answer was compared against each of the V, P and K scores for each question and the corresponding vata or pitta or kapha score was incremented based on the answer.
At the end, each subject had a vata score, a pitta score and a kapha score. These scores too were tabulated in another npz file which formed the output of the script. The output was of the following format.

	Subject	Vatta_score	Pitta_Score	Kapha_score
0	S017	1	6	1
1	S018	2	3	3
2	S020	2	4	3
3	S022	2	3	4
4	S029	3	3	1
...	...	...	...	...
135	S206	3	3	2
136	S207	1	4	2

This output was fed into the machine learning model against the audio parameters as input. The model worked to generate a relationship between those audio parameters and the vata scores, pitta scores and kapha scores of each subject in question.

RESULTS

Currently, the audio recording and responses to the questionnaire were collected from 100 subjects. The Table 1 provides the details of the subjects.

Table 1: Details of the subjects involved in data collection.

Total Subjects Male subjects Female subjects Mean and standard deviation of the subjects.

DISCUSSION

From the studies on Shabda Pariksha of Ayurveda, it was observed that the amount of three doshas (Vata, Pitta, and Kapha) could be predicted based on voice quality. The literature study also revealed that different personality traits are associated with the corresponding Tridoshas of the individuals. In this project, a study will be carried out to experimentally assert the relationship between the voice samples and the Tridoshas. Also, the project aims to extend the Shabda Pariksha to predict personality traits from voice samples. The project aims to obtain the correlation between voice features and the amount of Tridoshas, and the personality traits to assert the relationships. Further, machine learning models will be trained to predict the amount of Tridoshas and personality traits from the voice samples of an unseen subject.

To carry out such study a database providing the voice samples, the corresponding amount of three doshas of the person, and the corresponding personality traits is required. However, no such database from a large number of subjects is currently available. In the first phase of the project, a significant effort is put into collecting such a dataset. To ascertain good quality voice recordings the recording was done at a well-established studio setup with an anechoic chamber at the Indian Institute of Information Technology Dharwad. Currently, data from 100 subjects was collected from the students, research scholars, faculty, and teaching and non-teaching staff. Consent was taken from each of the subjects before recording. As data collection is a time-consuming task, we have distributed our work among ourselves equally to get a proper dataset of audio of people belonging to different backgrounds.

In the next phase of the internship, the correlation between acoustic features (such as pitch, intensity, and spectral characteristics) and recorded personal traits will be studied to derive some conclusions. We will also be collecting additional data, and training machine learning models to obtain an automatic predictor which can finally predict the amount of doshas and attributes based on any given new voice sample.

REFERENCES
1. Jaiswal YS, Williams LL. A glimpse of Ayurveda - The forgotten history and principles of Indian traditional medicine. J Tradit Complement Med. 2016 Feb 28;7(1):pp. 50-53. doi: 10.1016/j.jtcme.2016.02.002. PMID: 28053888; PMCID: PMC5198827.

2. Pandit Kashinatha Shastri, Charaka Samhita of Agnivesa, Hindi commentary Part-I Rogbhishakji vimanadhyaya,chapter 8.verse no.94 Varanasi:Chaukhambha Bharati Academy, pp.771.

3. Vaidya Laxmipati Sastri, Yogaratnakara with vidyotani Hindi commentary, Varanasi Chaukhambha Prakashana; Reprint Edition 2017, pp. 1 – 5.

4. Shastri L (2002) 7th (Edn.), Yoga Ratnakara of yaychandrashekhar, Shabda Pariksha. Chaukhamba Publications, Varanasi pp: 13.

5. Travis FT, Wallace RK. Dosha brain-types: A neural model of individual differences. J Ayurveda Integr Med. 2015 Oct-Dec;6(4):280-5. doi: 10.4103/0975-9476.172385.

6. Mills PJ, Peterson CT, Wilson KL, Pung MA, Patel S, Weiss L, Kshirsagar SG, Tanzi RE, Chopra D. Relationships among classifications of ayurvedic medicine diagnostics for imbalances and western measures of psychological states: An exploratory study. J Ayurveda Integr Med. 2019 Jul-Sep;10(3):198-202. doi: 10.1016/j.jaim.2018.02.001.

7. P. Venkata Giri Kumar, Sudheer Deshpande, H.R. Nagendra, Traditional practices and recent advances in Nadi Pariksha: A comprehensive review, Journal of Ayurveda and Integrative Medicine, Volume 10, Issue 4, 2019, Pages 308-315

8. Subhash Waghe, Critical Evaluation of Ayurvedic Shabda Pariksha (Auscultation) In Diagnosis of Diseases WSR to Computerized Sound Analysis, Journal of Natural & Ayurvedic Medicine, 2022, Volume 6, Issue 2, pp. 1-7.

9. Adesh Kumar and Sunil R. Khandare. A review article of Ashtavidha Pariksha with special reference on visham jwara. European Journal of Molecular & Clinical Medicine, vol. 7, issue 11, 2021, pp. 6299-6311.

10. Subhash Waghe, Critical Evaluation of Ayurvedic Shabda Pariksha (Auscultation) in Diagnosis of Diseases WSR to Disease Kasa (Cough), Science Journal of Clinical Medicine. Volume 11, Issue 3, September 2022 , pp. 57-62. doi: 10.11648/j.sjcm.20221103.11

APPENDIX I
Please attach consent form as given by Chinmay Sir.

APPENDIX II

List of 36 Hind phonemes recorded for quantifying the pronunciation of basic phoneme sounds.

☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7
☐8 ☐9 ☐10 ☐11
☐12 ☐13 ☐14 ☐15
☐16 ☐17 ☐18 ☐19 ☐20
☐21 ☐22 ☐23 ☐24 ☐25
☐26   ☐27 ☐28 ☐29
☐30 ☐31 ☐32 ☐33
☐34   ☐35 ☐36

APPENDIX III
The process of data collection involves two questionnaires in order to record the vatta, pita, kapha doshas of the subject. Certain questions in the questionnaires are as follows:

- Would you describe your appetite as irregular, normal or always hungry?
- Would you say your facial skin is dry or oily? Is it rough or smooth?
- Do you characterise your hair structure as straight, wavy or curly?
- Are your teeth uneven or well-formed?
- Would you say your speed of learning a concept is less, average or more?
- Do you tend to remember things short term, long term or somewhere in between?
- How much do you overthink while implementing a given task? How quickly do you implement it, on a scale of 1(least) to 3(most)?
- How lazy are you while implementing a task, on a scale of 1(least) to 3(most)?
- Can your sleep be characterized as light and disturbed, moderate and regular or deep and heavy?
- If I say you worry alot, how much will you agree with that statement on a scale of 1(least) to 3(most)?
- What time do you usually wake up?
- How often do you dream?
- When you're thinking about things do you tend to think more about the past, present or future?
- How often do you clean your room?
- Do you easily forgive people, does it take time or do you never forgive a person?
- How frequently do you get angry?
- Which type of movies are you most interested in? Rate your interest in action films, horro films, comedy films, sci-fi films, romantic films and suspense/thriller films on a scale of 1(strongly dislike) to 5(strongly like).
- Rate your interest in dancing, drawing/painting and travelling on a scale of 1(strongly dislike) to 5(strongly like).
- Rate your interest in drinking hot drinks and cool drinks on a scale of 1(strongly dislike) to 5(strongly like).
- Rate your interest in Mathematics related subjects on a scale of 1(strongly dislike) to 5(strongly like).
- Which type of music do you enjoy most?
- Which category of colours do you find most appealing? Rate the following colours on a scale of 1(strongly dislike) to 5(strongly like): red, yellow, green, blue, white, black and orange.
- How frequently do you face the following health issues, on a scale of 1(very rarely) to 5(very frequently): headaches, cough and cold, stomach related problems, joint related problems and back related problems.

Most of these characteristics about a person are a direct influence of the vatta, pitha, kafha doshas. Additionally, these traits aid the dataset we intend to create.