The K-Nearest Neighbour (kNN)-aided Predictions of Aquifer Fluoride Levels in the Rahole Area, Dadaab sub-County

Meshack Owira Amimo[1] and Dr K.S.S. Rakesh[2]

[1]*Research Scholar, IIC University of Technology, Cambodia*
[2]*CEO, Gradxs, India*

[1]bmoamimo@gmail.com
[2]kssrakesh@gmail.com

***Abstract*—Fluoride levels have been a challenge in the Merti aquifer, especially those that are located on the distal and peripheral fringes of the Merti. Some areas in these two zones have acceptable levels of fluorides whereas others do not have acceptable levels. Bone fluorosis has been a public health concern in the border areas and this defect has been blamed on suspected anomalous fluoride levels of some old wells.**

Rahole area is located in the distal Merti and is therefore a candidate for anomalous levels with a relatively high probability. Since the study area is a new settlement, the hydrogeological assessment incorporated a study into the expected levels of fluorides so that the borehole is sunk with this knowledge in mind-the expected levels of fluoride. The levels may not be suitable for humans but still be favorable for livestock. To achieve this, a Machine Learning algorithm, K-Nearest Neighbor, was used. The results generated prove over ninety percent accuracy levels of predictions. The study area is found to be free of fluoride levels deemed harmful to human life.

***Keywords*—KNN, K-Nearest Neighbors, Fluoride levels, Merti aquifer.**

I. INTRODUCTION

The project is situated 165 kms away from Garissa Township on the **northwestern flanks.** The road is all weather all the way, traversing **Garissa-Dadaab-Hagadera-Hammey stretch.** Already there is a borehole in the area, but this is a**t Damajaley,** managed by an existing water user committee, but is just very far away (20 kilometers) from the centre which has to subsist on water bowsers.

Efforts are being made to respond to the impacts of global Climate Change, in terms of trying to woo the community from hundred percent reliance on Nomadic lifestyle, and green-house farming is one way to go.

The water in the township alone cannot serve the livestock and human populations targeted in this predominantly grazing land. The water is meant to serve a population of at least **500** persons, and livestock approximating 15000 shoats, camels and cows.

A discharge of approximately 10-27$\mathbf{m^3}$per hour should suffice these intended purposes. This amount is highly possible from the proposed borehole, given its location within **the Merti Aquifer** fringes.

Also, from the values of resistivity generated, the amount of minerals in the aquifer is expected range from to be fair and tolerable to almost nil salinity. The numerous anthills dotting the countryside are a testimony to the tolerability of the subsurface minerals to animal life.

II. PROJECT LOCATION

A. Location

The project area lies in North eastern province within **Dadaab subcounty** and is located on the Northeastern sides of the Garissa Township. The area is defined foretasted at an altitude of approximately **138m**above sea level. Oblique dipping sediments litter the terrain alongside some zero degree dipping units of Miocene Pliocene sediments which are located in the area.

B. Nature of the Project

This is a communial water supply project meant to enhance the rapid settlement of the local population. Already there are boreholes in the far-flung **center, Damajaley** and also at **Hammey,** but these are not enough to accommodate the needs factored earlier.

The Center has a daily requirement of approximately 20-cubic metres of water to address the needs foretasted.

The borehole will be developed upon drilling completion and should be preferably encased with **Johnson screens or** bituminous coated plain casings / slotted casings. Once the productivity of the borehole has been determined, a suitable submersible pump will be installed to pump water into the proposed storage tanks. The schematic design and the detailed itemization for the proposed borehole shall be the subject of phase two work for the planning and designing unit, but will be predicated on the borehole performance in terms of aquifer yields and recharge. In case the yield will be too low for a submersible motor powered pumpage, a hand pump or windmill driven pump system is suitable as well.

C. Project Ownership

The proposed site is a public facility owned by the registered Rahole **Community rural Water Project**. At an appropriate stage, the community will be trained on management related issues pertaining to sanitation and operations and maintenance.

D. Site alternative and proposed action

There is a second site in the same area, located some 100m away from the first, but it is not necessary as the site earmarked for drilling is promising to be sufficiently productive. Moreover, the resistivities generated promise **good quality water.**

III. HYDROGEOLOGY

A. Geology and Stratigraphy

The topography is undulating dotted with several anthills which are clayey rich, and support vegetations that comprise mainly thorny shrubs, undergrowths and acacia family trees.

The geology is defined by red to light toned sandy clayey sediments, the Jurassic Corallite formation, which overlies the carbonates – namely corallites, aragonitic sediments and calcite. The sandy clayey species are mainly the Mariakani Sandstones series-mainly to a depth of 80m. From 80m onwards, we have fine to coarse grained sandstones dominating the geology, alongside silts and gravel. The Miocene/Pliocene story terminates at the Basement contact zone beyond 300m bgl.

The Jurassic limestone carbonates are fairly fractured and possess water at the shallow depths, though highly mineralized, via the fractures and karstification veins. Water also forms at the contact points between the sandstones and the sedimentary shalestones encountered at great depths.

Groundwater in the upper sediments shall enjoy annual precipitation recharge through direct infiltration, while the deep-seated zones shall be recharged via regional flow aided by the karstification channels and plate tectonics in the Jurassic – cretaceous period. Evapo transpiration rates of up to 3,000mm per annum over shadow the annual rains of up to 400mm per annum.

B. Physiography

The area stands at an average altitude of **138m metres above sea level** within a gently dipping terrain punctuated with several ant hills and flood plains both on the south eastern and north western flanks.

IV. HYDROLOGY AND STRUCTURAL GEOLOGY

A. Recharge Mechanisms within the Aquifer Systems

Rahole area lies within the Merti aquifer. The recharge is a function of the flow in the Merti beds which leverages on the favorable porosity and transmissivity levels of the aquifers located in the subsurface. The sediments along which water flows are both homogeneous and heterogeneous. The geology of the sediments in Rahole is such that the aquifers are admitting heterogeneous flow into its systems since the resistivity values indicate sandstones and gravels as part of the geology. One can deduce flow velocity of recharge from aquifer temperatures which have been anomalous in the neighborhoods of Rahole. It is easy to deduce that the aquifer recharge flow feeding the area emanates primarily from the Dadaab area. Some small pockets of flow is from Welmerer and Alinjugur areas as well.

Owing to this heterogeneity, velocities of flow is higher in some parts of the flow traverse and lower in others, making the aquifer temperatures vary significantly. With this in mind, it is expected that Rahole aquifer temperatures may be as high as Hammey and Amuma boreholes.

B. Drainage

Owing to the relative flat nature of the terrain, there is flood rampancy. The semi-permanent civil structures on the ground to stand the risk of destruction added to the occasional loss of lives for both livestock and human persons. Most of the housing units are constructed through shrubs and dry acacia trees locally available, lightening the task of evacuation in the event of impending flood disasters.

C. Climate

The project area falls within zone 7 of the classification of climatic/ecological zones of Africa, that is to say arid to semi-arid with temperatures averaging 30 to 34 degrees per day and occasioning evapo transpiration rates of up to 3000mm per annum. The rainfall average falls well below 500mm per year.

V. GEOPHYSICS

A. Introduction

In order to determine the Projects Area's hydrostragraphy and aquifer suitability, a total of 3No. Vertical Electrical Soundings were undertaken using an ABEM SAS 4000B Terrameter. Schlumberger arrays were used so that current electrode spreads of up to 250m against potential spreads of between 5m and 25m were employed to conduct the surveys. Copper electrodes were used for the potentials, while steel iron electrodes were used for the currents. The best site had its data analysed for drilling preparations.

B. Analyzed Geophysical Data

The analyzed Geophysical data generated via the Schlumberger configurations are appended hereunder.

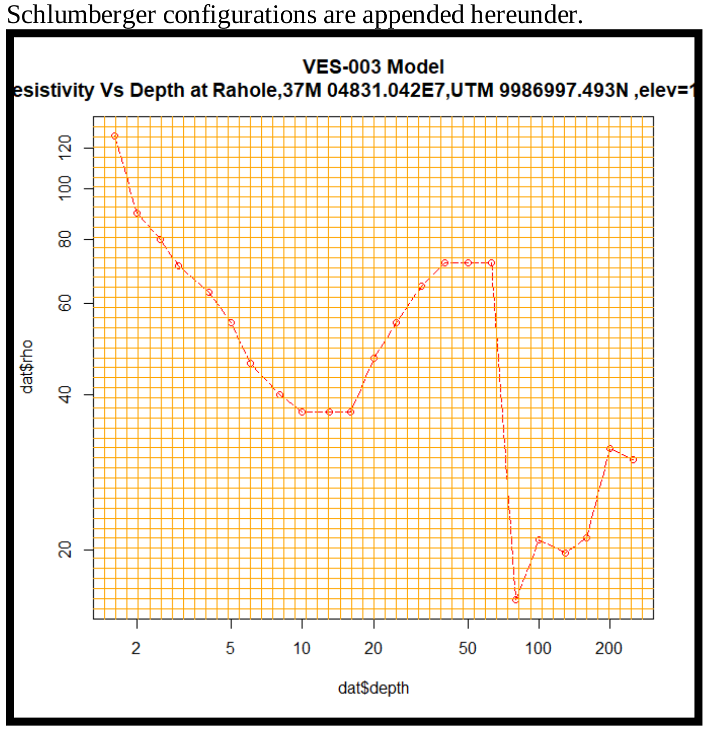


Figure 1: The geophysical graph of 1D showing resistivity against depths of the best spot selected for drilling

TABLE I
ANALYZED DATA FOR THE BEST SITE-UNDER AN ACACIA

RESISTIVITY CURVE NO.	FORMATION DEPTH INTERVAL (M)	RESISTIVITY (OHM.M)	EXPECTED GEOLOGICAL FORMATION
R-001/2020	0-1.6	127	Red Topsoils
	1.6-10	37	Subsoils
The third site located under	10-20	47	Saturated Loams/sands
an acacia.	20-40	72	Fine sands
The site located in a zone	40-60	72	Sandstones
where soils are reddish	60-100	21.1	Sandstones/clays
	100-200	31.5	Sandstones & Clays
	200-275	29.9	Medium & coarse sands
	Over 275	Infinity	Sandstones

MAP 1-TOPOGRAPHIC MAP SHOWING RAHOLE AREA

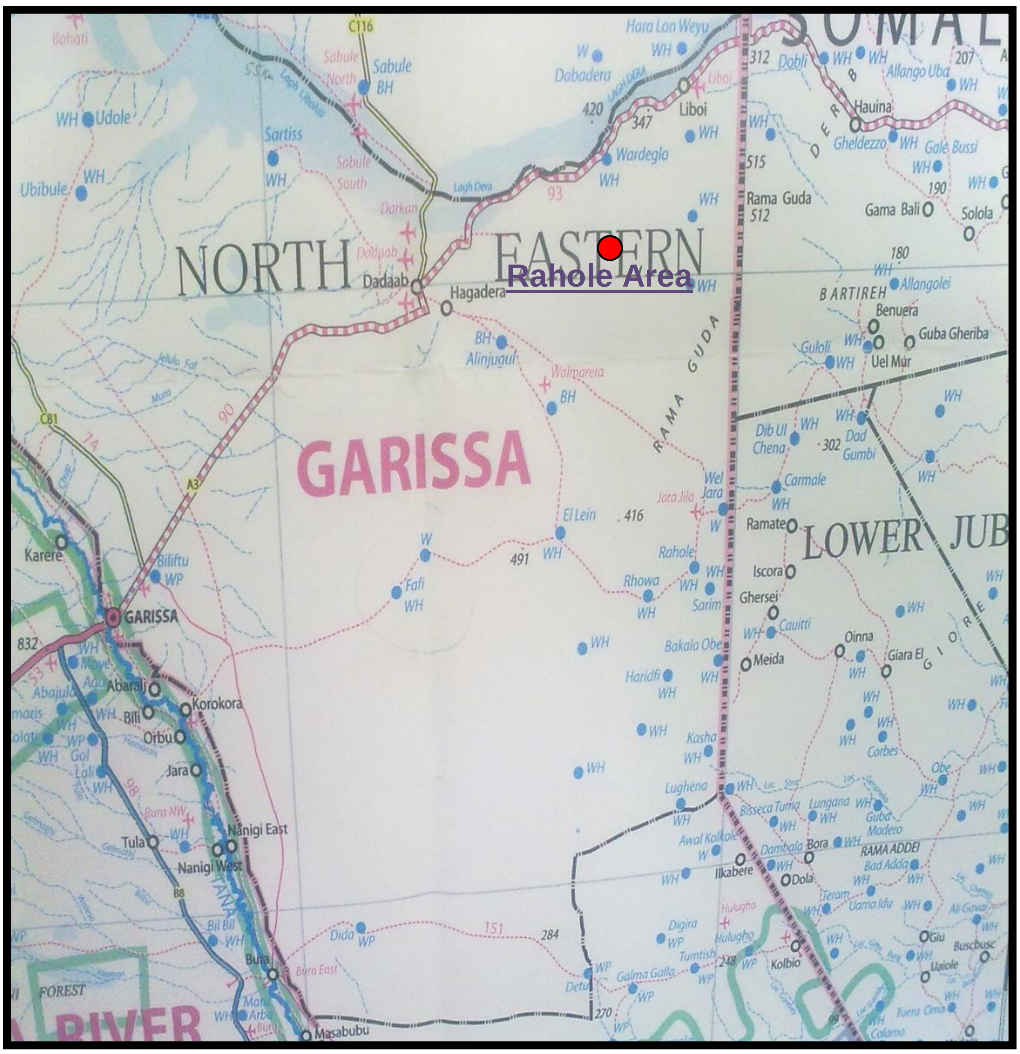


Figure 2: Map Showing the Location of the Proposed Borehole of Rahole

VI. LITERATURE REVIEW

Fluoride levels has been an issue in the study area especially the Amuma-Sidax Gose areas which are located in the border zones with Somalia. They both neighbor the Rahole area. Bone malformations in children and weak teeth, from fluorosis-related problems, have been a public health challenge, necessitating careful planning during water development to ensure we don't strike high-fluoride aquifers.

Since fluoride has been such an issue, it is imperative that fluoride levels be investigated and predicted before the well is sunk.

In a 2021 study by Liu et al, some three hundred and twenty four borehole water samples were subjected to statistical analysis using piper diagrams and saturation index methods, as well as the GIS software, ArcGIS. The study establishes that groundwater fluoridelevels were negatively correlated with calcium levels in a relationship deemed to be of statistical significance. The fluoride ions were established to bear positive correlation with levels of pH, bicarbonates and sodium ions. In this study, the significant parameters or factors that came out as having an effect on groundwater fluoride concentrations in the study area included the dissolution thermodynamics of fluorites and calcite components of the aquifers.

Another study was undertaken in West Africa in 2019 by Zango et al. In the study, some eighty eight groundwater samples were analysed. In North eastern Ghana. The origin of the aquifers waters, the aquifer chemistry and the geospatial distribution of the waters were analysed. Aquifer Boron and Fluoride levels were also assessed. The fluorides and Boron levels were noted to be anomalous in zones that were actively contaminated via geogenic sources. The control factors that determined their levels in the groundwaters mapped were found to be mainly hydrochemical factors. This included the aridity of the study areas, general climatic factors, the alkaline chemistry of the aquifers, weathering geological thermodynamics, among others.

A study in Pameano area of Argentina in 2016 also yielded results offering deep in sight into geological factors determining fluoride mineralization in ground water (Zabala et al, 2016). The study employed multivariate statistics to map levels of fluoride and arsenic. It established increase of Fluoride levels to the north-eastern portion of study area which was the mapped direction of flow of groundwater. This may be significant in that the Merti also realizes this same increase in levels of fluorides in the flow direction, towards Somalia.

The major hydro-geochemical processes actively controlling both fluoride and arsenic geo-distribution in the study area were found out to be the anion-cation exchange dynamics, with the sodium-ion release and calcium-ion uptake, carbonate-dissolution and the pH-level rise.

A study in India in 2018 by Srivastava et al aimed at establishing the various sources of groundwater aquifer contamination. The study employed both graphical and statistical methods.

The geo-spatial spreads of aquifer nitrates in the study indicated such a primary contribution of agricultural fertilizers, to the observed anomalous levels of fluorides, so that the study concluded that fluoride-mineralization in aquifers occurred in the aquifer systems via the dissolution of fluoride-enriched geologic mineral compounds. As a solution to anomalous levels of fluorides, the study stressed on the need to reduce the excessive usage of fertilizers so that the reduced intake of fertilizers by groundwater systems would allow the aquifers systems to rid itself of excessive fluoride via recharge.

According to Ahmed et al (2020), the fluoride levels associated with granitic aquifers were found to be primarily due to dissolution dynamics, during the rock-water interactions. The study also noted that the high variability observed in the aquifer fluoride levels arose from the fact that fluorine-enriched mineral compounds were not uniformly-redistributed in the parent-rock. Study concludes that the aquifer fluoride-levels is the result of geo-chemical interventions. This may also include the role of anthropogenic, since pesticides and fertilizers are a significant factor in the aquifer contamination by fluorides.

Luo et al (2018) undertook a geochemical mapping of the hydrogeologic systems in the Yungchen Basin of China. In the study, both Hydrogeochemical and stable isotope analyses and geochemical modeling were undertaken to map out the primary processes which controlled the groundwater quality as well as the fluoride concentration levels. The major outputs of the case study: that the anion-cation exchange dynamics and salt effects formed the primary controls on the geo-enrichment of fluoride in Yungchen aquifers in the area by reducing the activity of Ca^{2+}/F^{-} ratio in groundwater.

Moreover, anthropogenic-alluded pollution via pesticide and fertilizer-use, as well as industrial waste effluent-discharges was also a major source of fluoride-concentration in the local aquifers with the other sign cant cause being the desorption of fluoride from mineral or organic-matter surfaces.

VII. STUDY METHODOLOGY IN ESTIMATING FLUORIDE LEVELS

To achieve this, a surfer map of the Merti aquifer was generated using existing and newly generated GPS data. The longitudes and latitudes of the project area were then inserted, and the approximate levels of fluoride shown.

Deshmukh et al (2016) used GIS and Surfer to map groundwater aquifer contamination and water quality and derived reliable results, as was Al Naeem et al (2019)

In this study, the csv excel file data of hydrochemical and GIS data were prepared and analysed using R software. The estimation of Fluoride was thus two-fold:

a) The GIS geospatial spread.
b) The ML analysis of the fluoride levels

. Rahole area is seen plotted on the green color, which is from the key given under the Model representing class 2.

This is very good water quality meaning it ranges from 1mg/L upto 1.5mg/L.
The model prepared was with these assumptions.
1-excellent fluoride levels (0.1 to 1.0mg/L)

2-very good fluoride range (1-1.5 mg/L)

3-poor to very poor fluoride range (over 1.5 mg/L levels).

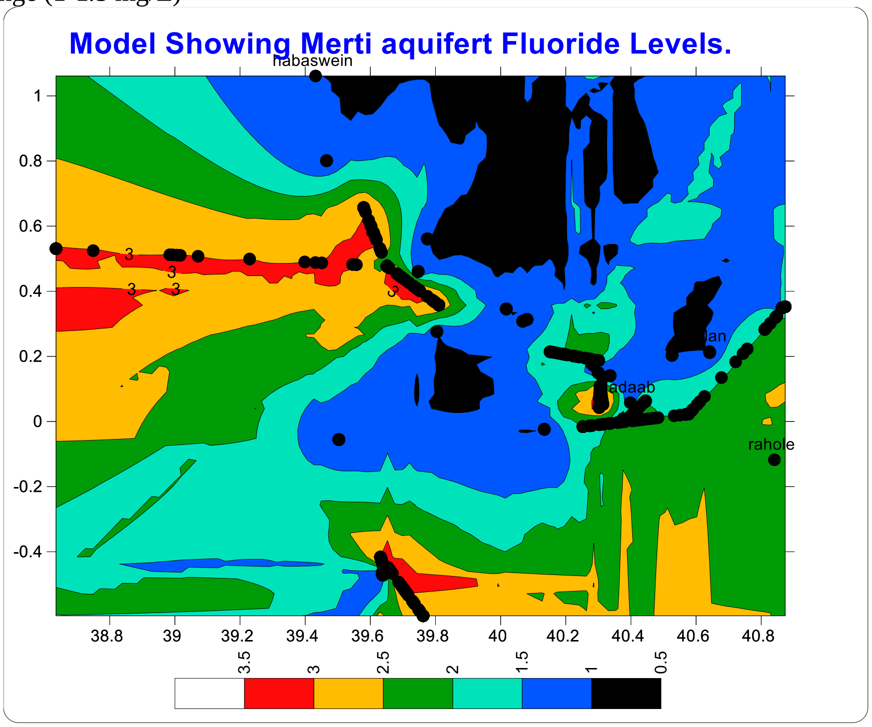


Fig 3: Model of Fluoride Mineralization in the Wells of Merti Groundwater Systems.

The model of depths to aquifer in the Merti is shown hereunder- location of Rahole area, and the fact that there exist no significant relationship between fluorides and Depths in Rahole comes out.

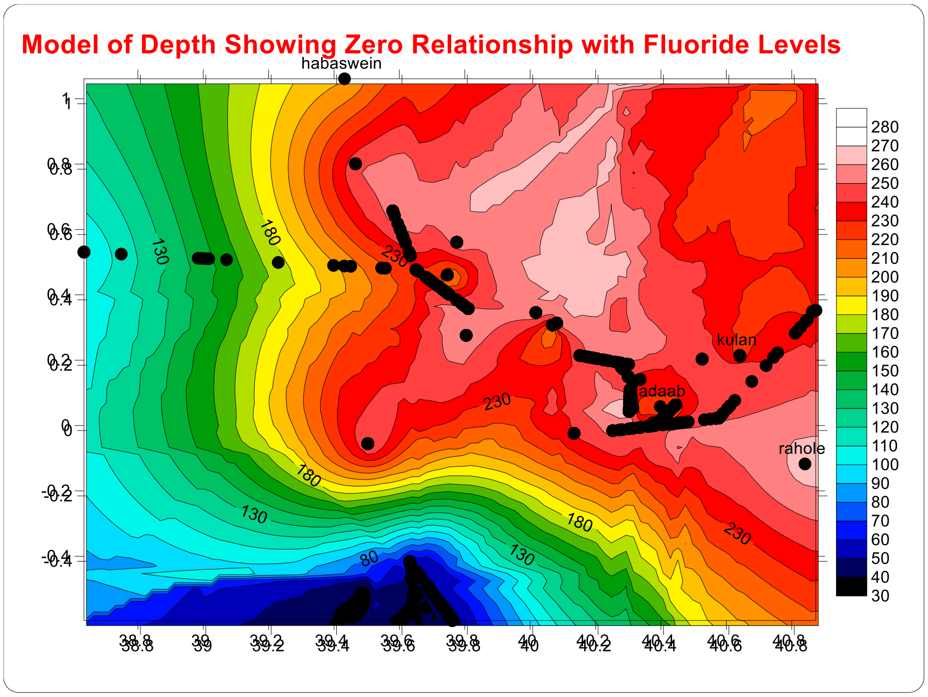


Figure 4- GIS Model of Depth Variations. The map models would appear more or less the same for both depths and fluorides were there to be a significant correlation between the two variable.

TABLE II
COORDINATES OF RAHOLE AREA WHICH WAS SURVEYED AND PROPOSED FOR DRILLING

Longitude(longtd)	Latitude (lattd)	Elevation (elev)	Proposed Depth (in meters)
40.84045	-0.11756	138	265

VIII. PREDICTION USING K-NEAREST NEIGHBOR IN R

This is an algorithm which employs the Hamming, Euclidian distance or the Minkowski algorithms to generate a prediction. In the present study, the Euclidean distance method was used.
The algorithm is simple to use and is illustrated hereunder.

Euclidean $$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

A. *How the algorithm works*

When used in tasks of classifying variable rows for different categories, the K-nearest neighbor algorithm primarily functions in such a manner as to be the equivalent of determining class on the basis of a ‘simple majority vote’. This implies settling for a number of rows, K, in such a way that this K is always an odd number. The number of rows with more attributes is assigned the new row whose class is unknown (names are arbitrary and do not represent actual places in this example table), as shall be illustrated hereunder:

	A	B	C	D	E	F
1	longtd	lattd	elev	aquiferDepth	classSodium	name
2	30.8	1.233	355	240	high	saretho
3	30.82	1.244	309	238	high	abakaile
4	30.953	1.237	345	300	high	kalala
5	30.789	1.251	321	302	high	fafiajin
6	30.68	1.254	297	312	high	modogashe
7	35.3	1.255	200	302	low	dadab
8	34.46	1.3	309	421	low	hagadera
9	33.6	1.556	321	399	low	alinjugur
10	32.6	1.332	265	298	low	welmerer
11	32.7	1.301	345	287	NA	sidaxGose

Figure 5: Excel Screenshot Showing the Classes of Fluoride

The tenth row has an NA in the **classSodium** column. The fictitious place is sidaxGose.

Longitude=32.7, Latitude= 1.301, Elevation= 345, And aquifer Depth=287.

All the places with names are areas with water and whose sodic levels have already been mapped. We went to the field and measured the coordinates and depths to expected aquifer at SidaxGose. We don't know the aquifer sodium levels. We don't know the sodic levels in the aquifer to be developed.

In the excel sheet, the Euclidian distance has been made to work in such a way that the values of the un-classified row is subtracted from each of the columns with known values, squared, added and then the summed value evaluated for its square root, getting the values of this Euclidean distance for all the sites as follows:

=((H17-A17)^2+(I17-B17)^2+(J17-C17)^2+(K17-D17)^2)^0.5

The operation above on excel sheet gives out the following table.

Table III

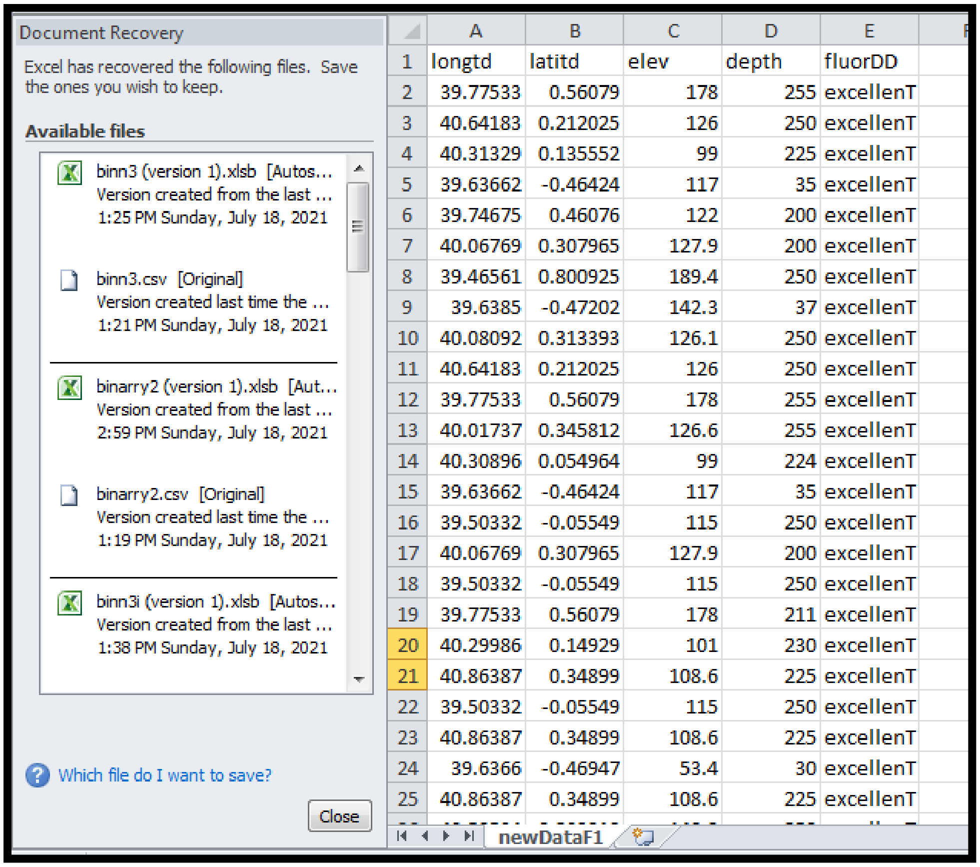
longtd	lattd	elev	aquiferDepth	classSodium	longtd	lattd	elev	aquiferDepth	eucldn Dist		
30.8	1.233	355	240	high	32.7	1.301	345	287	48.08965	high	3
30.82	1.244	309	238	high	32.7	1.301	345	287	60.83204	high	5
30.953	1.237	345	300	high	32.7	1.301	345	287	13.11702	high	1
30.789	1.251	321	302	high	32.7	1.301	345	287	28.36643	high	2
30.68	1.254	297	312	high	32.7	1.301	345	287	54.15794	high	4
35.3	1.255	200	302	low	32.7	1.301	345	287	145.797	low	10
34.46	1.3	309	421	low	32.7	1.301	345	287	138.7627	low	9
33.6	1.556	321	399	low	32.7	1.301	345	287	114.5464	low	8
32.6	1.332	265	298	low	32.7	1.301	345	287	80.75278	low	7
33.7	1.4556	300	332	low	32.7	1.301	345	287	63.64765	low	6

The ranking in red at the table where Euclidean distances have been calculated are all representing high sodium levels. The new class is thus of 'high 'levels of sodium. If two rows had been 'high' and a single row had been 'low', the class would still be 'high' for sodium

B. K-Nearest Neighbours in the present study

The K nearest neighbor algorithm was used on the data and predictions made on the expected fluorides in the Rahole area. R software was used to map the fluoride levels as shown in the discussion.

The screenshot of the data is shown hereunder:



	A	B	C	D	E
1	longtd	latitd	elev	depth	fluorDD
2	39.77533	0.56079	178	255	excellenT
3	40.64183	0.212025	126	250	excellenT
4	40.31329	0.135552	99	225	excellenT
5	39.63662	-0.46424	117	35	excellenT
6	39.74675	0.46076	122	200	excellenT
7	40.06769	0.307965	127.9	200	excellenT
8	39.46561	0.800925	189.4	250	excellenT
9	39.6385	-0.47202	142.3	37	excellenT
10	40.08092	0.313393	126.1	250	excellenT
11	40.64183	0.212025	126	250	excellenT
12	39.77533	0.56079	178	255	excellenT
13	40.01737	0.345812	126.6	255	excellenT
14	40.30896	0.054964	99	224	excellenT
15	39.63662	-0.46424	117	35	excellenT
16	39.50332	-0.05549	115	250	excellenT
17	40.06769	0.307965	127.9	200	excellenT
18	39.50332	-0.05549	115	250	excellenT
19	39.77533	0.56079	178	211	excellenT
20	40.29986	0.14929	101	230	excellenT
21	40.86387	0.34899	108.6	225	excellenT
22	39.50332	-0.05549	115	250	excellenT
23	40.86387	0.34899	108.6	225	excellenT
24	39.6366	-0.46947	53.4	30	excellenT
25	40.86387	0.34899	108.6	225	excellenT

Figure 6: Screenshot of Csv Data used In the Model.

The algorithm runs as shown hereunder:

```
dataX=read.csv("newDataF1.csv",header=T,na.strings="NA")

head(dataX)

attach(dataX)


# split into train (train) and test (test)
train_index <- sample(seq_len(nrow(dataX)), size = round(0.75*nrow(dataX)))
train <- dataX[train_index, ]
test <- dataX[-train_index, ]

# take a look at the sample
head(train)
head(test)

# save class from later
train_status = train$fluorDD
test_status = test$fluorDD

# exclude FluorDD from train and test dataset
train = train[-5]
test = test[-5]

# runs knn
library(class)
```

Figure 7: Screenshot of Algorithm of kNN Scripts as ran In R Software

The algorithm generates very reliable prediction accuracy as may be seen in the screen shot shown overleaf, predicting only three incorrect instances in a total of 217 rows. This means that the accuracy of this kNN model is:

Accuracy=100 %*(217-3)*1/217

Accuracy=100*214/217

Accuracy= 98.61751 percent

```
R Console

             | dataX_test_knn
 test_status |   excellenT | unsuitabble |       Vgood |   Row Total |
-------------|-------------|-------------|-------------|-------------|
   excellenT |         106 |           0 |           0 |         106 |
             |       1.000 |       0.000 |       0.000 |       0.488 |
             |       0.991 |       0.000 |       0.000 |             |
             |       0.488 |       0.000 |       0.000 |             |
-------------|-------------|-------------|-------------|-------------|
 unsuitabble |           1 |          59 |           1 |          61 |
             |       0.016 |       0.967 |       0.016 |       0.281 |
             |       0.009 |       0.983 |       0.020 |             |
             |       0.005 |       0.272 |       0.005 |             |
-------------|-------------|-------------|-------------|-------------|
       Vgood |           0 |           1 |          49 |          50 |
             |       0.000 |       0.020 |       0.980 |       0.230 |
             |       0.000 |       0.017 |       0.980 |             |
             |       0.000 |       0.005 |       0.226 |             |
-------------|-------------|-------------|-------------|-------------|
Column Total |         107 |          60 |          50 |         217 |
             |       0.493 |       0.276 |       0.230 |             |
-------------|-------------|-------------|-------------|-------------|
```

Figure 8: The screenshot showing the power of model predictive capability of kNN with the data for predicting fluoride levels using Machine Learning in R software

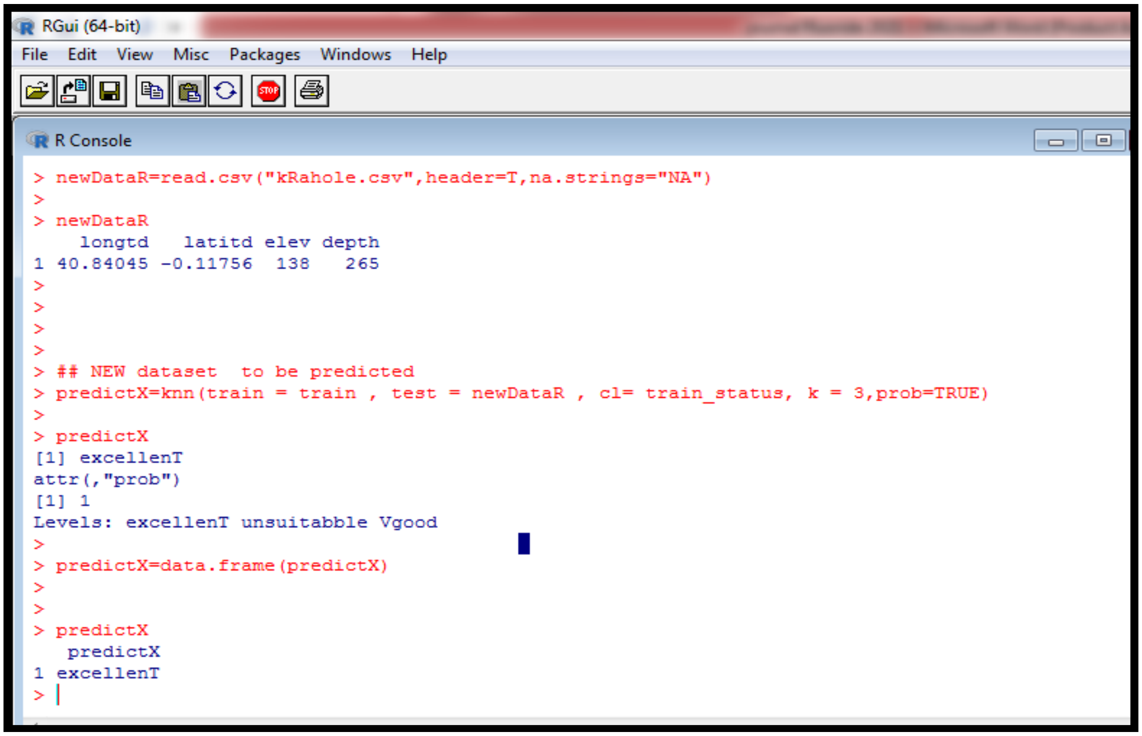


Figure 9- Showing the Results of kNN prediction which categorizes Rahole as having excellent fluoride levels-ranging from 0.1 to 1.0 mg/L. The GIS method had given it a category 2 (which implies 1, 0 to 1.5 mg/L) prediction. This is rated as very Good.

C. Summary Example with Application

An area has come up for groundwater development and the donor, Government of Kenya is concerned that the water to be drilled may be way too fluoride-mineralized for human use. The table hereunder shows the coordinates of the areas to be drilled.

TABLE III
LOCATION COORDINATED OF THE PROPOSED BOREHOLES SITES

Longitude(longtd)	Latitude (lattd)	Elevation (elev)	Proposed Depth (in meters)
40.63182	0.212025	165	255
39.77533	0.56081	179	225

We may now use the algorithm generated earlier to prove or disprove their fears. The predictive kNN algorithm in R was as thus:

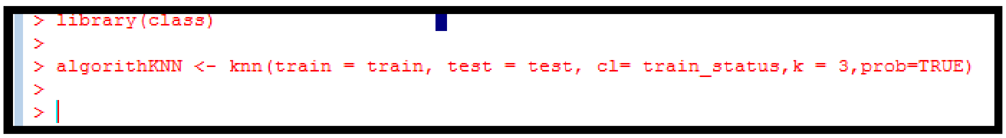


Figure 10: The Screenshot Showing the R GUI kNN Algorithm Calculator

We shall replace the ‘test=test’ portion of the algorithm with the newTestData as thus, then predict:

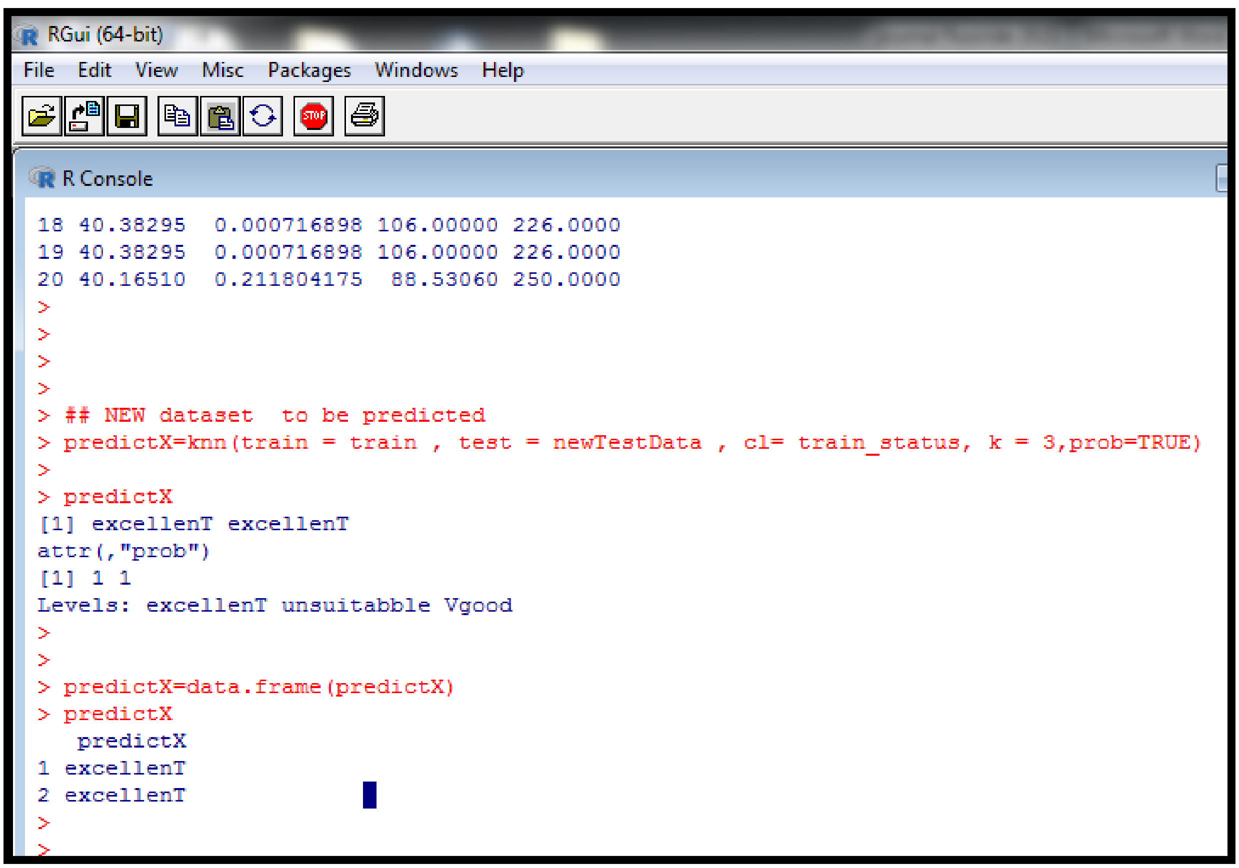


Figure 11: Predictions of new proposed project areas as excellent in terms of being fluoride free.

IX. RECOMMENDATIONS AND CONCLUSIONS

The present study indicates that using the GIS data, taken using the field GPS, in combination with the depths generated using the geoelectrical survey Terrameter was able to help predict fluoride levels for Rahole as very good using both statistical and cartographic softwares. The R machine learning algorithms, using kNN, predicted the Rahole area as excellently suited for water in terms of acceptable fluoride levels. The study thus recommends sinking of abstraction wells and development of water supply infrastructure to aid its distribution and use

Areas in the neighborhoods deemed suitable for future development by the government in Dadaab sub-county were also predicted for fluoride contamination using the same algorithms and the results were accurate up to 98.6 %.

X. REFERENCES

[1] Ahmed, S., & Sreedevi, P. D. (2020). Cyclic variation of fluoride contents with time in a granitic aquifer in semi-arid region. In *Hydrology and Water Resources* (pp. 199-210). CRC Press.

[2] Al Naeem, M. F. A., Yusoff, I., Ng, T. F., Maity, J. P., Alias, Y., May, R., & Alborsh, H. (2019). A study on the impact of anthropogenic and geogenic factors on groundwater salinization and seawater intrusion in Gaza coastal aquifer, Palestine: An integrated multi-techniques approach. *Journal of African Earth Sciences*, *156*, 75-93.

[3] Deshmukh, K. K., & Aher, S. P. (2016). Assessment of the impact of municipal solid waste on groundwater quality near the Sangamner City using GIS approach. *Water resources management*, *30*(7), 2425-2443.

[4] Liu, J., Peng, Y., Li, C., Gao, Z., & Chen, S. (2021). A characterization of groundwater fluoride, influencing factors and risk to human health in the southwest plain of Shandong Province, North China. *Ecotoxicology and Environmental Safety*, *207*, 111512.

[5] Luo, W., Gao, X., & Zhang, X. (2018). Geochemical processes controlling the groundwater chemistry and fluoride contamination in the Yuncheng Basin, China—An area with complex hydrogeochemical conditions. *PloS one*, *13*(7), e0199082.

[6] Srivastava, S. K., & Ramanathan, A. L. (2018). Geochemical assessment of fluoride enrichment and nitrate contamination in groundwater in hard-rock aquifer by using graphical and statistical methods. *Journal of Earth System Science*, *127*(7), 1-23.

[7] Zabala, M. E., Manzano, M., & Vives, L. (2016). Assessment of processes controlling the regional distribution of fluoride and arsenic in groundwater of the Pampeano Aquifer in the Del Azul Creek basin (Argentina). *Journal of hydrology*, *541*, 1067-1087.

[8] Zango, M. S., Sunkari, E. D., Abu, M., & Lermi, A. (2019). Hydrogeochemical controls and human health risk assessment of groundwater fluoride and boron in the semi-arid North East region of Ghana. *Journal of Geochemical Exploration*, *207*, 106363.