IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

Optimisation of Acid Yellow Dye Decolourisation using CNS Carbon: A Response Surface Methodology Approach with Exploratory Data Analysis in Jupyter Notebook

Vinithaa Raguram[1] and Rajalakshmi Padmanaban[2]
B. Tech Chemical Engineering, Botminds.ai, Chennai, India[1]
M. Tech Chemical Engineering, Anna University, Chennai, India [2]
Corresponding author – vini291990@gmail.com
rajalakshmiactau@gmail.com

Abstract: *The current research is focused on the need to develop an efficient adsorbent with low-cost effectiveness and high potentiality for removal of dye from waste water. Cashew nut shell, a solid waste generated in the cashew nut industry is taken as the adsorbent for the decolorization of acid yellow synthetic dye solution. Batch adsorption studies were carried out with the consideration of factors/variables such as dye concentration, adsorbent dosage and contact time. A comparative decolorization study has been done with both commercial activated carbon (CAC) and Cashew nut shell carbon (CNSC). The factors were optimized statistically using Central Composite Design (CCD) in response surface methodology. The characterization of the absorbent was analyzed in FTIR, SEM and XRD. In recent years, the use of interactive computing environments has significantly improved the efficiency of data analysis workflow. This study investigates the effectiveness using Jupyter notebooks in facilitating exploratory data analysis and reproducible research practices. Using the datasets, we conducted a comprehensive analysis involving data cleaning, visualization, and statistical exploration, leveraging Python libraries such as numpy, pandas, seaborn, and matplotlib*

Keywords: Acid yellow dye, Adsorption, Commercial Activated Carbon, Cashew nut Shell, Central composite design, Jupyter notebooks

I. INTRODUCTION

Synthetic dyes are used to cause immediate visible pollution in effluent water aside from contamination by interfering with light, which may retard photosynthesis and biota growth. They increase oxidation demand and lower vital oxygen levels in waters [1, 2]. Synthetic dyes are generally resistant to biodegradation and physiochemical techniques for their removal, such as adsorption, chemical oxidation, electro-coagulation and advance oxidation processes have been recently used extensively to comply with more and more stringent legislation regarding maximum allowable dye concentration in discharged waste water [3, 4]. Adsorption by activated carbon has been proven to be a successful physiochemical technique for the treatment of dye-laden wastewater mainly for its simplicity, efficiency and ease of implementation [5-7].

The utilization of alternative low-cost materials with high adsorption activity to solve environmental problems has received considerable attention over the recent years. Adsorbents from agriculture by-products are particularly advantageous due to their low cost and high availability as many other materials, for example, clays[8], cane waste[9], wood[10], cellulosic materials [11] and fish scale [12] have been tested as adsorbent on remediation of contaminated water.

DOI: 10.48175/IJARSCT-28706

The study to be reported deals with investigating the impact of supplementing mechanically conducted preliminary treatment which includes screening with adsorption on the treatability of tannery waste water containing acid yellow dye 100.

Acid Yellow dye 100 has the IUPAC name 4-Aminonaphthalene-1-sulfonic acid di azo, coupled with 2-Hydroxybenzoic acid and then translate into chromium complex. Its structure is Single azo, metal Complexes. Molecular formula is $C_{17}H_{10}N_2Na_2O_6S$ and Molecular weight is 416.32.

For the reduction of concentration of this dye to the environment, the natural waste material cashew nut shell generally known as CNS was used. This CNS has abundant carbon content when it is burnt, usually CNS is the waste material which used as feed instead of wood because of its ignition capacity. CNS has a thick, highly viscous liquid in it which is mostly used for engine ignition on proportions with other oil .

In this study, we have treated the acid yellow dye using CNS carbon and Activated carbon (purchased) to find the effectiveness of both carbons. Normally, activated carbons taken from market are meant to have high % of removal of concentration compared to any natural products, but due to the expensive nature of commercial activated carbon an alternate usage of CNS Carbon is preferred in this study.

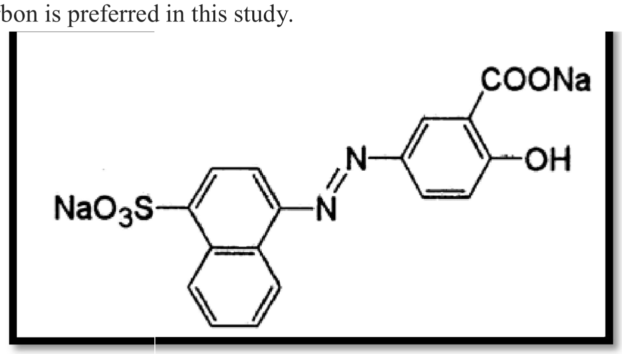


Figure 1. Structure of Acid Yellow 100 dye

II. MATERIALS AND METHODS

Preparation of Carbon from Cashew Shell nut

The cashew nut shell (CNS) was collected from local agriculture fields near Tamil nadu, India. The CNS was washed thoroughly 3–4 times in tap water to remove dust and other impurities adhered to it. It was then dried in a hot air oven at 60°C for 24 h. The Commercial activated carbon (CAC) was purchased from Lobachemie ltd, India. The cashew nut shell carbon (CNSC) was used as adsorbents and analytical techniques were performed.

Exploratory Data Analysis

Jupyter Notebook is an open-source web-based interactive computing platform that allows users to create and share documents containing live code, equations, visualizations, and narrative text. It supports various programming languages, most commonly Python, and is widely used in data science, machine learning, and scientific computing. In this study, Jupyter Notebook was used for exploratory data analysis (EDA), preprocessing, and visualization of the dataset. Its interactive nature facilitated iterative analysis, enabling real-time feedback on code execution and data outputs. The reproducibility and transparency of the computational workflow were enhanced by embedding code, results, and explanatory commentary in a single document.

Preparation of dye solution

The Acid Yellow 100 dye (molecular formula is $C_{17}H_{10}N_2Na_2O_6S$ and molecular weight is 416.32) was supplied by tannery. Double distilled water was used for the preparation of all the solutions with pH 6.5. A1000mg/L stock solution was prepared and 250,500,750mg/L solutions were prepared and the concentrations of the acid yellow solutions were obtained from standard calibrations curve.

Characterization of the Cashew Nut Shell Carbon (CNSC)

The CNSC were examined using FTIR spectrum, SEM - EDS analysis and X-Ray diffractometer. FTIR was employed to determine the presence of surface functional groups in the samples (Agilent Technologies). The crystal structure was characterized by X-Ray diffractometer (XRD) with Cu, Kα radiation (X Ray Pro powder method). The micro structure of the carbon was performed by Scanning Electron Microscope and the quantity of carbon was obtained by EDS analysis (High resolution field emission electron microscope with EDS, nano manipulation system).

Design of Experiments (DOE)- Response surface methodology:

Design of experiments is used to systematically investigate the process variables that influence the output quality with optimization. Response surface methodology (RSM) is a tool in DOE that is widely used to assess the relationship between one or more response variables and a set of quantitative experimental variables or factors. There are three steps in RSM (a) design and experiments, (b) response surface modeling through regression, and (c) optimization. Central composite design (CCD) is often recommended for optimizing the experimental variables in the study.

Independent process parameters can be represented in quantitative form using RSM as:

$$Y = f(X_1, X_2, X_3, X_4 \dots X_n) \quad (1)$$

where, Y represents the response (yield), f is the response function, ε is the experimental error, and X_1, X_2, X_3 … X_n are independent parameters. A surface, known as the response surface is

obtained by plotting the expected response of Y.

The response surface is obtained by plotting the expected response of Y. Thus, RSM aims at approximating f by a suitable lower ordered polynomial in some region of the independent process variables. If the response can be well-suited by a linear function of the independent variables [Prem Kumaret al.2014], the function (Eq. (1)) can be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \cdots + \beta_n X_n \pm \varepsilon \quad (2)$$

Experimental studies

Yellow dye adsorption experiments onto CNS and CCS from aqueous solution was performed in batch scale at pH 6.5 and 50 RPM .The experiments were conducted in 250 ml conical flasks containing 100ml soultion of dye solution with adsorbent concentration ranging form 1-2g/ per 100ml each, which was scaled up in g/l as shown in table 2. Interactive effects were studied between chosen variables (Adsorbent dosage, Dye Concentration and Contact time) on the maximum removal of dye.

Independent Variables	Design Variables	Range and levels		
		-1	**0**	**1**
Adsorbent dosage(g/l)	X_1	4	6	8
Dye concentration (mg/l)	X_2	50	75	100
Contact time (mins)	X_3	60	75	90

Table 1 : Level of independent variables and experimental range

Since, the independent variables is three in the present study, a full factorial CCD in this study consists of 8 factorial points, 6 axial points and 6 replicates at the centre with total of 20 sets of experiments under CCD with suitable combinations of independent variables chosen using RSM were calculated by the Eq. (3) and the details are presented in Table 3.

$$N = 2^{K-1} + 2^K + n_0 \quad (3)$$

Where N is the total number of experiments and k is the number of factors. The experiments were designed from CCD under RSM using MINITAB 16.0. Adsorbent dosage, Dye Concentration and Contact time were taken as independent variables and the carbon removal rate (%) was taken as dependent output response variable in this study. The experimental design matrices of three individual variables with regard to their uncoded and coded values are given in Table 3. Estimation of quality of fit of the model is done through coefficient determination and analysis of variances. A

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, August 2025 Impact Factor: 7.67

mathematical model in form of second-order polynomial Eq. (4) was used to fit the experimental results as a function of independent variables involving their interactions.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 \quad (4)$$

Where,
Y represents the dependent variable (Dye removal (%);
X_1, X_2, X_3represents the independent variable;
$\beta_0$represents the regressioncoefficient at center point;
$\beta_1, \beta_2, \beta_3$ represents the linear coefficients;
$\beta_{11}, \beta_{22}, \beta_{33}$representsthe quadratic coefficients; and $\beta_{12}, \beta_{13}, \beta_{23}$represents the second-order interaction coefficients.
The obtained regression model was assessed by taking the regression coefficient values, analysis of variance (ANOVA), P- and F-values. The coefficient of determination $R^2$was used to express the goodness of fit of the polynomial model equation. To identify the experimental design and to generate a regression model to find the optimum combinations considering the effects of linear, quadratic, and interactive effects on removal of dye, the statistical software package (MINITAB16.0) was used.

Run	X_1	X_2	X_3	Adsorbent Dosage (mg/l)	Dye Concentration (mg/l)	Contact time (min)
1	-1	-1	-1	4	50	60
2	1	-1	-1	8	50	60
3	-1	1	-1	4	100	60
4	1	1	-1	8	100	60
5	-1	-1	1	4	50	90
6	1	-1	1	8	50	90
7	-1	1	1	4	100	90
8	1	1	1	8	100	90
9	-1.6817	0	0	2.63	75	75
10	1.6817	0	0	9.36	75	75
11	0	-1.6817	0	6	32.95	75
12	0	1.6817	0	6	117.04	75
13	0	0	-1.6817	6	75	49.7
14	0	0	1.6817	6	75	100.22
15	0	0	0	6	75	75
16	0	0	0	6	75	75
17	0	0	0	6	75	75
18	0	0	0	6	75	75
19	0	0	0	6	75	75
20	0	0	0	6	75	75

Table 2: Full factorial design with code and real variables

III. RESULTS AND DISCUSSION

Characterization of the adsorbent

The surface characteristics of carbon samples can be seen by SEM photographs analyzed at 50µm, in which pores on the surface confirming the carbon. The availability of pores and internal surface is necessary for an effective adsorbent and percentage of carbon from EDS analysis are performed resulting in effective carbonaceous material.
Fourier Transform Infrared Spectroscopy provides the qualitative information about the functional groups of the surface. The absorption capacity of the carbon is also studied by this surface chemical structure, it was attributed at

DOI: 10.48175/IJARSCT-28706

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, August 2025 Impact Factor: 7.67

2283 cm^{-1} and 2210 cm^{-1} indicating alkyl C=C stretch bonds ,1238cm^{-1} and 1182.903 cm^{-1} indicating C-C bonds. Unsaturated compounds are attributed above 3000cm^{-1}.

The X-Ray diffractometer analysis are performed, the peaks are found around 2θ = 26° and 43°, the anode material used was Cu having broad peaks similar to crystalline carbonaceous structure such as graphite (JCPDS).

Regression Equation Development

The dye removal results of various design variables (Adsorbent dosage, Dye concentration, and Contact time) are shown in the Table4. The experimental results of maximum dye removal (%) from the calculated regression coefficients shown in Table 4 and Table 6 are fitted to second order polynomial Eq.4. The Eq.4 with regression coefficient for removal of dye by CNSC and CAC is given in the Equation 5 and Equation 6 respectively

$$Y = 92.1646 + 10.9846X_1 - 7.7342X_2 + 3.7066X_3 - 4.6386X_1^2 - 3.5508X_2^2 - 13.4527X_3^2 - 0.49X_1X_2 + 1.015150X_1X_3 + 1.0125X_2X_3 \quad (5)$$

$$Y = 99.2692 + 0.3008X_1 + 0.18X_2 - 0.3497X_3 - 0.1381X_1^2 - 0.2608X_2^2 - 0.0098X_3^2 + 0.0778X_1X_2 + 0.1647X_1X_3 + 0.5047X_2X_3 \quad (6)$$

The predicted output responses by CCD were correlated with experimental data with help of regression coefficient R^2. Furthermore, the proportion of variation in the outcome is given by the model and is denoted by $R^2$and adjusted R^2 values, where R denotes the variation in the observed responses. The measure of fit of the model is also given by the value of R^2, and the model is compared with different independent variables by the adjusted R^2 values. The R^2 evaluates correlation between experimental and predicted values [premkumar].

The R^2 value for CNSC and CAC were 0.91 and 0.87 obtained is closer to 1, which indicates the perfect fit of the data presented in Table 4 and 5. The obtained R^2 value indicates that 91% and 87% of variability in dye removal by both the adsorbents (CNSC and CAC) could be explained by the model.

Proximate and ultimate analysis

Proximate analysis of CNS and its char

Proximate analysis of CNS and its Char for determination of moisture content and volatile matter,

	Moisture content	Volatile matter
CNS	10-11%	65-70%
Char	5-6%	25-30%

Ultimate analysis of CNS and its char

The Ultimate analysis of CNS and its Char was carried out in order to determine its carbon, hydrogen, oxygen and nitrogen percentage.

	Carbon	Hydrogen	Oxygen	Nitrogen
Percentage	60-62%	6 to 7%	29-31%	0.7-0.75%

Ordinary Least Squares (OLS) Regression

OLS is a statistical method used to model the relationship between one or more independent variables (predictors) and a dependent variable (outcome). It is the most common form of linear regression.

OLS estimates the best-fitting line (or hyperplane, in higher dimensions) through the data by minimizing the sum of the squared differences between the actual values and the predicted values—these differences are called residuals.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 2, August 2025 Impact Factor: 7.67

Run	X_1	X_2	X_3	Ad. Dos (mg/l)	Dye Con. (mg/l)	Contact Time (mins)	CNSC-1	CAC	CNSC-1 Pred.	CAC Expt.
1	-1	-1	-1	4	50	60	64.8000	99.760	65.1029	99.4766
2	1	-1	-1	8	50	60	85.1500	99.980	86.0222	99.5933
3	-1	1	-1	4	100	60	56.7000	98.810	48.5896	98.6716
4	1	1	-1	8	100	60	72.1400	99.400	67.5489	99.0992
5	-1	-1	1	4	50	90	70.7000	97.322	68.4610	97.4382
6	1	-1	1	8	50	90	92.1600	98.260	93.4403	98.2139
7	-1	1	1	4	100	90	63.7000	98.450	55.9978	98.6522
8	1	1	1	8	100	90	86.1500	99.640	79.0171	99.7388
9	-1.68	0	0	2.63	75	75	53.3100	98.400	60.5707	98.3726
10	1.6817	0	0	9.36	75	75	95.1200	99.096	97.5185	99.3845
11	0	-1.6817	0	6	32.95	75	98.5501	97.961	95.1286	98.2288
12	0	1.6817	0	6	117.04	75	56.0336	98.841	69.1142	98.8342
13	0	0	-1.6817	6	75	49.7	44.3200	99.259	47.8809	99.8296
14	0	0	1.6817	6	75	100.22	54.2500	98.963	60.3483	98.6534
15	0	0	0	6	75	75	92.4408	99.120	92.1646	99.2692
16	0	0	0	6	75	75	92.4408	99.340	92.1646	99.2692
17	0	0	0	6	75	75	92.4408	99.300	92.1646	99.2692
18	0	0	0	6	75	75	92.4408	99.300	92.1646	99.2692
19	0	0	0	6	75	75	92.4408	99.300	92.1646	99.2692
20	0	0	0	6	75	75	92.4408	99.300	92.1646	99.2692

Table 3 : Full factorial design with experimental and predicted results of adsorbents

Fischer's test (F-test) is used to calculate the ratio of mean square of regression (MRR) to error (MRe).The smaller the extent of the F-value yields more significant is the corresponding coefficient. Moreover, if the model has high degree of adequacy for predicting the experimental results, the computed F-value from the model should be greater than the tabulated F-value.From Table 4&6, it can be noted that the obtained F-value for dye removal by ACS and CCS is 12.3 and 7.48, which is greater than the tabulated F-value.

The calculated regression coefficient for dye removal is given in Table 5 and Table 7 for ACS and CCS together with their corresponding p-value and T-value. From the table 5, the coefficient for single effect of Adsorbent dosage (β_1), and Dye concentration(β_2), square effects of adsorbent dosage (β_{11}) and contact time (β_{33}) are significant (p<0.050).Similarly, From the table 7, the coefficient for single effect of Adsorbent dosage (β_1), and Contact time (β_3), interactive effects of dye concentration and contact time (β_{23}) are significant (p<0.050).

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	9	5491.55	491.55	10.17	2.30	.000
Residual Error	10	496.17	496.17	49.62		
Total	19	5987.72				

R-Sq = 91.71% ; R-Sq(adj) = 84.26%

Table 4 Analysis of variance CNSC

Term	Coefficient	SE Coefficient	T	P
β_0	92.1646	2.873	32.081	0.000
β_1	10.9846	1.906	5.763	0.000
β_2	-7.7341	1.906	-4.058	0.002
β_3	3.7066	1.906	1.945	0.080

β_{11}	-4.6386	1.856	2.500	0.031
β_{22}	-3.5508	1.856	1.914	0.085
β_{33}	-13.4527	1.856	7.250	0.000
β_{12}	-0.4900	2.490	-0.197	0.848
β_{13}	1.0150	2.490	0.408	0.692
β_{23}	1.0125	2.490	0.407	0.693

Table 5 Estimated Regression coefficient for CNSC

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	9	6.82594	6.82594	0.75844	7.48	0.002
Residual Error	10	1.01405	1.01405	0.10140		
		7.8399				

R-Sq = 87.07% ; R-Sq(adj) = 75.42%

Table 6 Analysis of Variance for CAC

Term	Coefficient	SE Coefficient	T	P
β_0	99.2692	0.12988	764.340	0.000
β_1	0.3008	0.08617	3.491	0.006
β_2	0.1800	0.08617	2.089	0.063
β_3	-0.3497	0.08617	-4.058	0.002
β_{11}	-0.1381	0.08388	-1.647	0.131
β_{22}	-0.2608	0.08388	-3.109	0.011
β_{33}	-0.0098	0.08388	-0.117	0.909
β_{12}	0.0778	0.11259	0.691	0.506
β_{13}	0.1647	0.11259	1.463	0.174
β_{23}	0.5047	0.11259	4.483	0.001

Table 7 Estimated Regression Coefficients for CAC

Main Effect Plots

These plots are used to calculate data means when several factors are involved in the process and to compare magnitudes of marginal means. The points in the plot represent the means of the response variables at various levels of each factor, with a reference line drawn at the grand mean of the response data. The MEP of ACS adsorbent on dye removal with effect of process variables is shown in figure 2. It can be noted that maximum dye removal was of 95.12%, 98.5%, and 85.76% were observed at 9.36(g/l), 32.95(mg/l), and 75mins. Furthermore, it can also be observed that lower dye concentration (32.955mg/l) yielded maximum removal (98.5%) and higher dye concentration (117.045mg/l) resulted in less removal (56.03%). Similarly, in 75mins of Contact time 85.75% of dye removal has been achieved, beyond that desorption process occurs.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

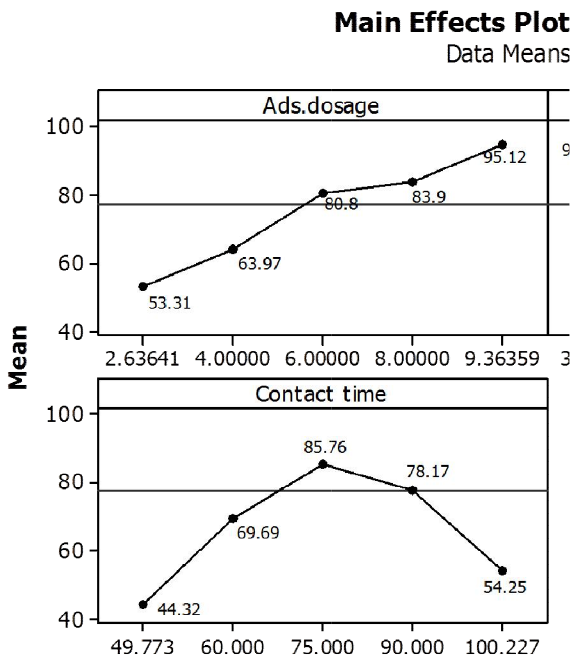


Figure 2 : Main effect plots on Adsorbent dosage, Concentration and Contact time for ACS

In the case of CAC which is shown in figure 3. It is observed at adsorbent dosage 8(g/l); Dye concentration of 75mg/l with contact time 60mins resulted in 99.13%, 99.32 % and 99.48% of dye removal .On the other hand, at very high concentrations of Adsorbent dosage (9.36mg/l) and Dye concentration (117.045mg/l) , there is decrease in the dye removal was observed.

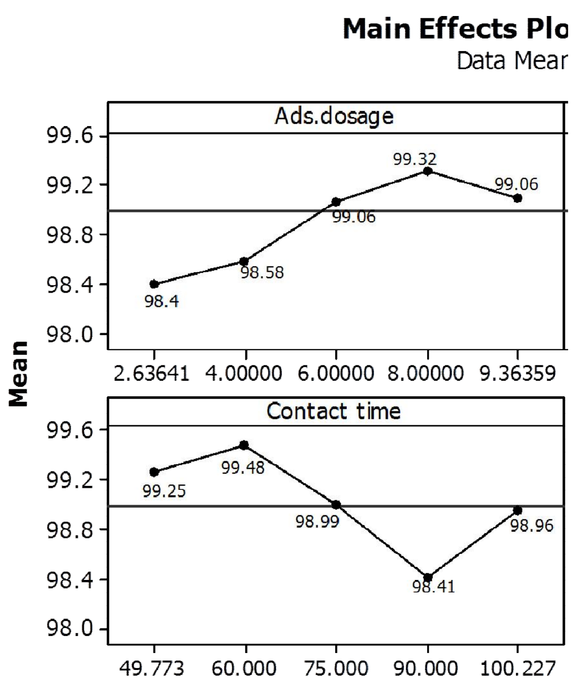


Figure 3 : Main effect plots on Adsorbent dosage, Concentration and Contact time on CAC

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, August 2025 Impact Factor: 7.67

From the MEP, it can be noted that at high adsorbent dosage, >90% of dye removal has been observed, this may be due to the fact that as the increase in availability of surface sites which is resulted from the increase dosage and conglomeration of the adsorbent [13]. With high concentration of the dye both adsorbents exhibited less removal which may be due to the lack of available sites in the surface of the adsorbent. Also, it can be noted that as the contact time increases, viz., after 75mins the average mean data of dye removal of both adsorbents did not show much significant in dye removal rate [14].

Response surface plots

The response surface plot for the carbon prepared form activated cashew shell (ACS) on dye removal is shown in the figure 4. It can be noted from the figure **a**. that increases in adsorbent dosage increases dye removal. However, as the dye concentration increases, after 75mg/l of dye concentration there was decrease in the reduction of dye removal respective of increase in adsorbent dosage. Similar trends were observed in figure **b**. On the other hand, it can be noted from the figure **c,** that higher dye concentration showed less removal at lower contact time, but as the contact time increases, increase in dye removal was observed till 75mins after which desorption starts. This shows that increase in dye concentration decrease the rate of external diffusion and enhancing stationary phase collision of adsorbed particles takes place as a result sorption is reduced.

Similarly, the response surface plots for CAC as shown in figure 5. It can be observed form figure **a,** that as the higher adsorbent dosage showed slight lower trend than initial dosage. Meantime, from figure **b,** it can be noted that at higher dosage, there was not much significant change in the dye removal at increasing contact time. However, at lower adsorbent dosage the dye removal was less as the contact time increases. This condition was similarly observed in figure **c** . This shows that at initial contact maximum sites in the adsorbent is filled at higher adsorbent dosage. At same time, initial adsorbent dosage there was decrease in dye removal as the contact time increase may be due to the fact that increase in the movement of adsorbed particles causes decrease in the film resistance to mass transfer in the adsorbent [15].

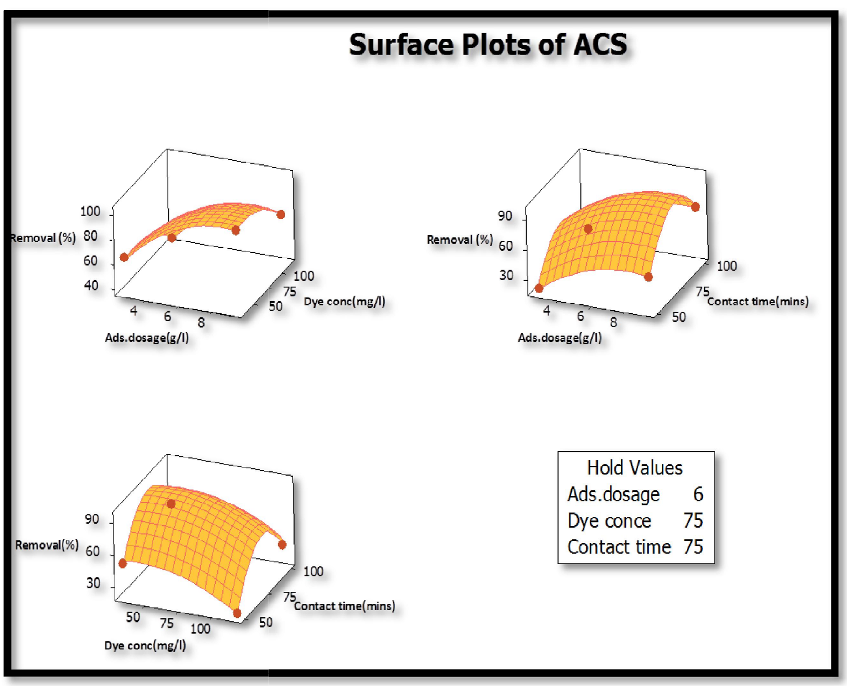


Figure 4: Response Surface plots of CNSC

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

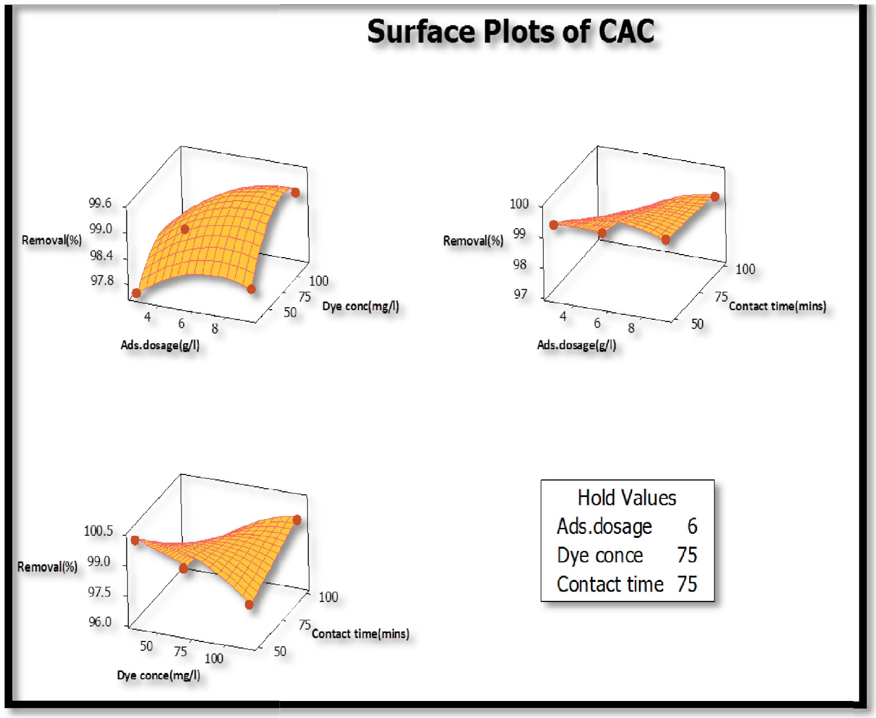


Figure 5: Response Surface plots of CAC

Exploratory Data Analysis

Program to import libraries

```
import pandas as pd
import statsmodels.api as sm
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

EDA for CNSC Experiment Vs Prediction

Data frame creation and OLS regression

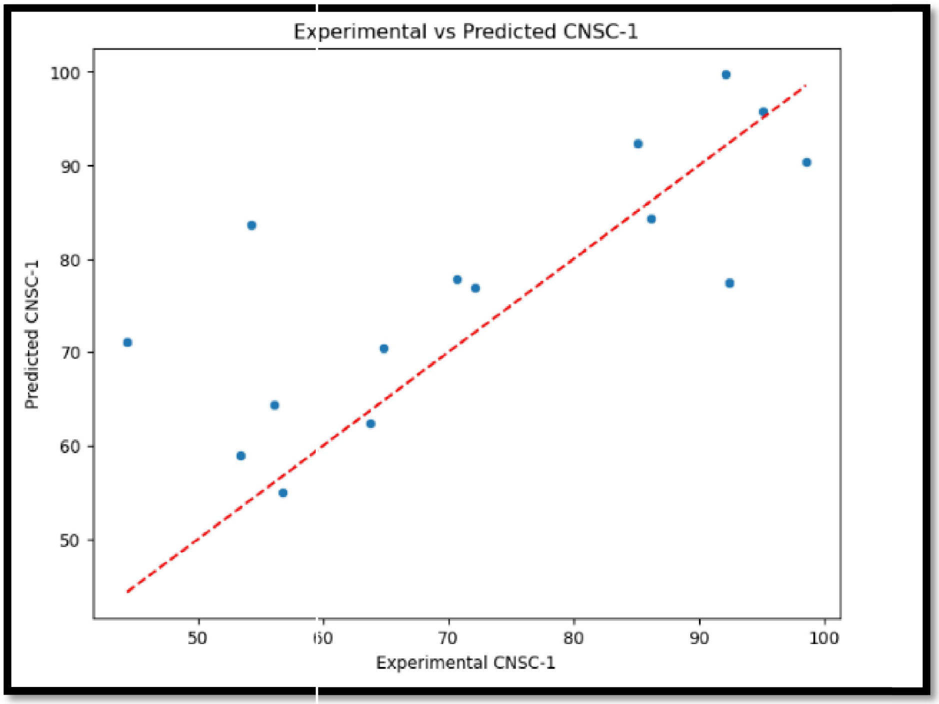
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

```python
data = {
    'X1': [-1, 1, -1, 1, -1, 1, -1, 1, -1.68, 1.6817, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
    'X2': [-1, -1, 1, 1, -1, -1, 1, 1, 0, 0, -1.6817, 1.6817, 0, 0, 0, 0, 0, 0, 0, 0],
    'X3': [-1, -1, -1, -1, 1, 1, 1, 1, 0, 0, 0, 0, -1.6817, 1.6817, 0, 0, 0, 0, 0, 0],
    'Ad.Dos(mg/l)': [4, 8, 4, 8, 4, 8, 4, 8, 2.63, 9.36, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6],
    'Dye Con.(mg/l)': [50, 50, 100, 100, 50, 50, 100, 100, 75, 75, 32.95, 117.04, 75, 75, 75, 75, 75, 75, 75, 75],
    'Contact time(mins)': [60, 60, 60, 60, 90, 90, 90, 90, 75, 75, 75, 75, 75, 75, 75, 75, 75, 75, 75, 75],
    'CNSC-1 Exp.': [64.8, 85.15, 56.7, 72.14, 70.7, 92.16, 63.7, 86.15, 53.31, 95.12, 98.5501, 56.0336, 44.32, 54.25, 92.44, 92.44, 92.44, 92.44, 92.44,
    'CAC Exp.': [99.76, 99.98, 98.81, 99.4, 97.32, 98.26, 98.45, 99.64, 98.4, 99.096, 97.961, 98.841, 99.259, 98.963, 99.12, 99.34, 99.3, 99.3, 99.3, 99
    'CNSC-1 Pred.': [65.1029, 86.0222, 48.5896, 67.5489, 68.4610, 93.4403, 55.9978, 79.0171, 60.5707, 97.5185, 95.1286, 69.1142, 47.8809, 60.3483, 92.16
    'CAC Pred.': [99.4766, 99.5933, 98.6716, 99.0992, 97.4382, 98.2139, 98.6522, 99.7388, 98.3726, 99.3845, 98.2288, 98.8342, 99.8296, 98.6534, 99.2692,
}

df = pd.DataFrame(data)

# Independent variables (X1, X2, X3)
X = df[['X1', 'X2', 'X3']]
X = sm.add_constant(X)  # Adds a constant term for the intercept

# Dependent variable for CNSC-1 (Experimental vs Predicted)
y_cns = df['CNSC-1 Exp.']

# Regression model for CNSC-1
model_cns = sm.OLS(y_cns, X).fit()

# Predicted CNSC-1 values from the model
df['CNSC-1 Pred. Model'] = model_cns.predict(X)

# Model summary
print("CNSC-1 Model Summary:")
print(model_cns.summary())

# Evaluate the model for CNSC-1 (R-squared, MAE, MSE, RMSE)
mae_cns = mean_absolute_error(y_cns, df['CNSC-1 Pred. Model'])
mse_cns = mean_squared_error(y_cns, df['CNSC-1 Pred. Model'])
rmse_cns = np.sqrt(mse_cns)
r2_cns = r2_score(y_cns, df['CNSC-1 Pred. Model'])

print(f"CNSC-1 Model Evaluation:\nMAE: {mae_cns}\nMSE: {mse_cns}\nRMSE: {rmse_cns}\nR-squared: {r2_cns}")

# Plot the experimental vs predicted for CNSC-1
plt.figure(figsize=(8, 6))
sns.scatterplot(x=y_cns, y=df['CNSC-1 Pred. Model'])
plt.plot([y_cns.min(), y_cns.max()], [y_cns.min(), y_cns.max()], color='red', linestyle='--')
plt.xlabel('Experimental CNSC-1')
plt.ylabel('Predicted CNSC-1')
plt.title('Experimental vs Predicted CNSC-1')
plt.show()
```

Output Model Summary

```
CNSC-1 Model Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:            CNSC-1 Exp.   R-squared:                       0.443
Model:                            OLS   Adj. R-squared:                  0.338
Method:                 Least Squares   F-statistic:                     4.241
Date:                Sun, 20 Apr 2025   Prob (F-statistic):             0.0220
Time:                        16:42:24   Log-Likelihood:                -79.545
No. Observations:                  20   AIC:                             167.1
Df Residuals:                      16   BIC:                             171.1
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         77.3853      3.228     23.969      0.000      70.541      84.229
X1            10.9865      3.908      2.811      0.013       2.702      19.271
X2            -7.7342      3.907     -1.980      0.065     -16.017       0.548
X3             3.7067      3.907      0.949      0.357      -4.576      11.989
==============================================================================
Omnibus:                        1.890   Durbin-Watson:                   0.889
Prob(Omnibus):                  0.389   Jarque-Bera (JB):                1.218
Skew:                          -0.600   Prob(JB):                        0.544
Kurtosis:                       2.860   Cond. No.                         1.21
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
CNSC-1 Model Evaluation:
MAE: 10.336337074577624
MSE: 166.7702260258126
RMSE: 12.913954701245185
R-squared: 0.4429453228467476
```

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67



EDA for CAC Experiment Vs Prediction

```
y_cac = df['CAC Exp.']
model_cac = sm.OLS(y_cac, X).fit()
df['CAC Pred. Model'] = model_cac.predict(X)

# Evaluate the model for CAC
mae_cac = mean_absolute_error(y_cac, df['CAC Pred. Model'])
mse_cac = mean_squared_error(y_cac, df['CAC Pred. Model'])
rmse_cac = np.sqrt(mse_cac)
r2_cac = r2_score(y_cac, df['CAC Pred. Model'])

print("CAC Model Summary:")
print(model_cac.summary())

print(f"CAC Model Evaluation:\nMAE: {mae_cac}\nMSE: {mse_cac}\nRMSE: {rmse_cac}\nR-squared: {r2_cac}")

# Plot the experimental vs predicted for CAC
plt.figure(figsize=(8, 6))
sns.scatterplot(x=y_cac, y=df['CAC Pred. Model'])
plt.plot([y_cac.min(), y_cac.max()], [y_cac.min(), y_cac.max()], color='red', linestyle='--')
plt.xlabel('Experimental CAC')
plt.ylabel('Predicted CAC')
plt.title('Experimental vs Predicted CAC')
plt.show()
```

IJARSCT

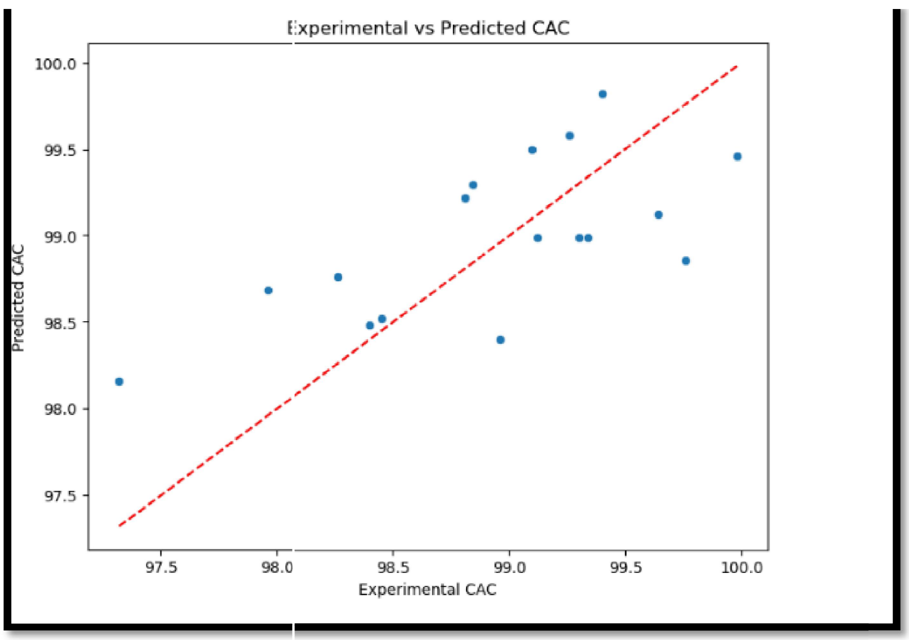
International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 2, August 2025 Impact Factor: 7.67

Output Model Summary

```
CAC Model Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:                CAC Exp.   R-squared:                       0.427
Model:                             OLS   Adj. R-squared:                  0.320
Method:                  Least Squares   F-statistic:                     3.977
Date:                 Sun, 20 Apr 2025   Prob (F-statistic):             0.0271
Time:                         16:45:55   Log-Likelihood:                -13.451
No. Observations:                   20   AIC:                             34.90
Df Residuals:                       16   BIC:                             38.88
Df Model:                            3
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         98.9900      0.119    835.234      0.000      98.739      99.241
X1             0.3010      0.143      2.099      0.052      -0.003       0.605
X2             0.1801      0.143      1.256      0.227      -0.124       0.484
X3            -0.3499      0.143     -2.439      0.027      -0.654      -0.046
==============================================================================
Omnibus:                         1.850   Durbin-Watson:                   0.779
Prob(Omnibus):                   0.396   Jarque-Bera (JB):                0.993
Skew:                           -0.037   Prob(JB):                        0.609
Kurtosis:                        1.911   Cond. No.                         1.21
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
CAC Model Evaluation:
MAE: 0.4220838132283767
MSE: 0.22474309566663261
RMSE: 0.47407077073642984
R-squared: 0.42716298009134124
```



IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 2, August 2025 Impact Factor: 7.67

IV. CONCLUSION

The objective is to use the efficacy of low-cost adsorbent over commercial adsorbent for tannery dye removal from aqueous solutions. In this study the effects of variables, such as adsorbent dosage, dye concentration and contact time on dye removal by CNSC and CAC powder using CCD under RSM with statistical analysis were performed. From the results, the dye removal by CNSC was 98.5% attained at adsorbent dosage 6g/l, dye concentration 32.95mg/l with contact time 75mins. Similarly for ACS, 99.980% of dye removal was achieved at adsorbent dosage of 8g/l , dye concentration of 50mg/l with contact time 60mins for CAS. The regression coefficient (R^2) value of 0.91 and 0.87 for CNSC and CAC, shows that experiments are statistically significant. The OLS regression analysis of the experimental data reveals that both CNSC-1 and CAC adsorbents are effective in removing dye from aqueous solutions, but their performance characteristics differ notably. CNSC-1 exhibits a strong dependency on process variables, particularly adsorbent dosage (X1) and contact time (X3), with higher values leading to significantly improved adsorption. The efficiency for CNSC-1 ranges more widely, indicating that careful optimization is necessary to achieve maximum performance. In contrast, CAC demonstrates consistently high adsorption efficiency (above 97%) across all experimental conditions, indicating a more stable and robust behavior with minimal sensitivity to variations in dye concentration, dosage, or contact time. The regression models for both materials show a good fit, with CAC having a slightly higher predictive accuracy ($R^2 \approx 0.99$). Overall, CAC is recommended for applications requiring high and stable dye removal efficiency under varying conditions, while CNSC-1 may be a suitable and cost-effective alternative when operating parameters can be tightly controlled and optimized

REFERENCES

[1]. V. Prigione, G.C. Varese, L. Casieri, C.F. Marchisio, Biosorption of simulated dyed effluents by inactivated fungal biomasses, Bioresource Technology,3559–3567, 99 (2008).

[2]. A. Ozer,G.Dursun, Removal ofmethylene blue fromaqueous solution by dehydrated wheat bran carbon, Journal of Hazardous Materials 262–269, 146 (2007).

[3]. X.S. Wang, Y. Zhoub, Y. Jiang, C. Sun, The removal of basic dyes from aqueous solutions using agricultural by-products, Journal of Hazardous Materials, 374–385, 157(2008).

[4]. B. Bestani, N. Benderdouche, B. Benstaali, A. Addou, Adsorption of Methylene Blue and iodine from aqueous solutions by a desert Salsola vermiculata species, Bioresource Technology, 8441–8444, 99 (2008).

[5]. K.R. Ramakrishna, T. Viyaraghavan, Dye removal using low cost adsorbent, Water Science Technology, 189–196, 36 (1997).

[6]. V. Meyer, F.H.H. Carlsson, R.A. Oellemann, Decolourisation of textile effluent using a low cost natural adsorbent material, Water Science Technology,1205,16 (1992).

[7]. H. Lata, V.K. Garg, R.K. Gupta, Removal of a basic dye from aqueous solution by adsorption using parthenium hyster, Dyes and pigments,653-658 , 74(2007).

[8]. L.C.A. Oliveira, R.V. Rios, J.D. Fabris, K. Sapag, V. Garg, R.M. Lago, Clay-iron oxide magnetic composite for the adsorption of contaminants in water, Appl. Clay Sci. 1–9, 803 (2002).

[9]. M. Nassar, Y.H. Magdy, Removal of different basic dyes from aqueous solution by adsorption on palm-fruit bunch particles, J. Chem. Eng., 223–229.

[10]. M. Asfour, M.M. Nassar, O.A. Fadali, M.S. EL-Geundi, Equilibrium studies on adsorption of basic dye on hardwood, J. Chem. Teuchn. Biotech. 257–263, 50(1991).

[11]. G. McKay, S.J. Allen, Single resistance mass transfer models for the adsorption of dyes on peat, J. Separ. Process Technol., 1–7, 4 (1983).

[12]. J.F. Villanueva-Espinosa, M. Hern´andez-Esparza, F.A. Ruiz-Trevi˜no, Adsortive properties of fish scales of Oreochromis Niloticus (Mojarra Tilapia) for metallic ion removal from waste water, Ind. Eng. Chem. Res. 3563–3569, 40 (2001).

[13]. Kannan N. and Karuppasamy K , Low cost adsorbents for the removal of phenyl Acetic acid from aqueous solution, *Indian J. Environ. Protec.,* , 683-690,18(1998).

[14]. N. Vedaraman, S. Shamshath Begum &S. V. Srinivasan, "Response surface methodology for decolourisation of leather dye using ozonation in a packed bed reactor"Clean Tech Environ Policy, 616,(2013).

[15]. Lee, C., Low, K., & Chung, L,.Removal of some organic dyes by hexane-extracted spent bleaching earth, Journal of Chemical Technology and Biotechnology,93–99, 69(1997).

[16]. João Felipe Pimentel, Leonardo Murta, Vanessa Braganholo , Juliana Freire.Understanding and improving the quality and reproducibility of Jupyter notebooks

APPENDIX

SEM & EDS Analysis

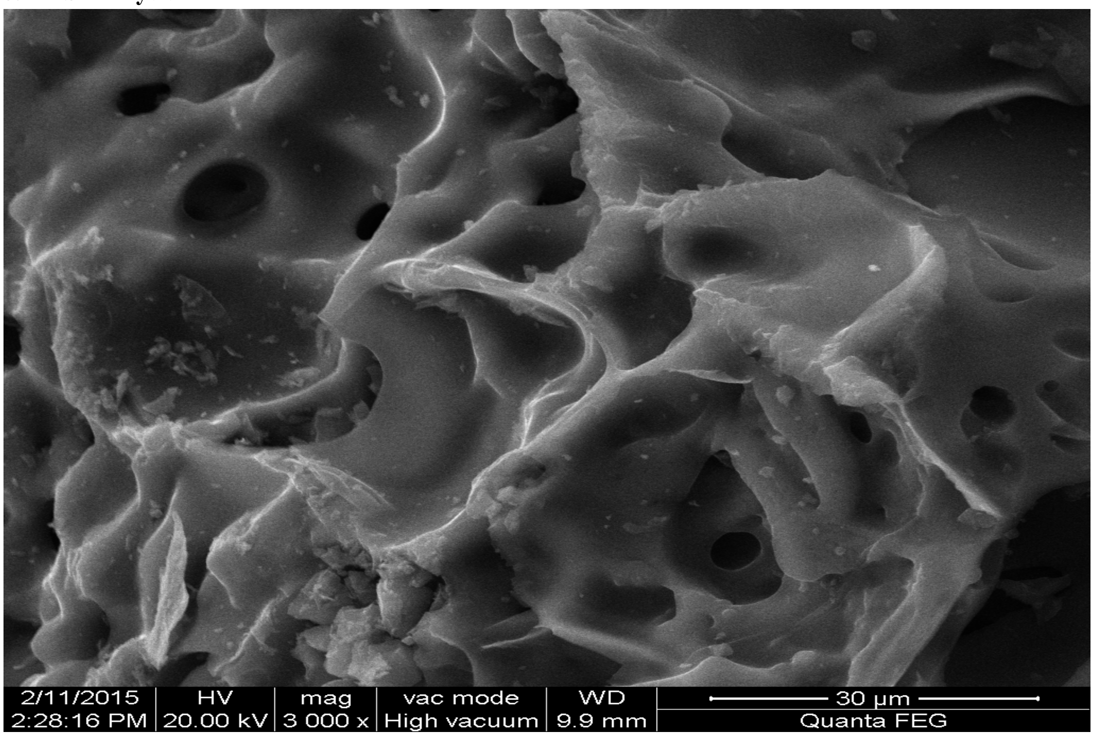


Figure 7 SEM analysis of CNSC before treatment

EDS Chart

Spectrum: training

El AN Series unn. C norm. C Atom. C Error (1 Sigma)

[wt.%] [wt.%] [at.%] [wt.%]

C 6 K-series 100.00 100.00 100.00 17.21

Total: 100.00 100.00 100.00

DOI: 10.48175/IJARSCT-28706

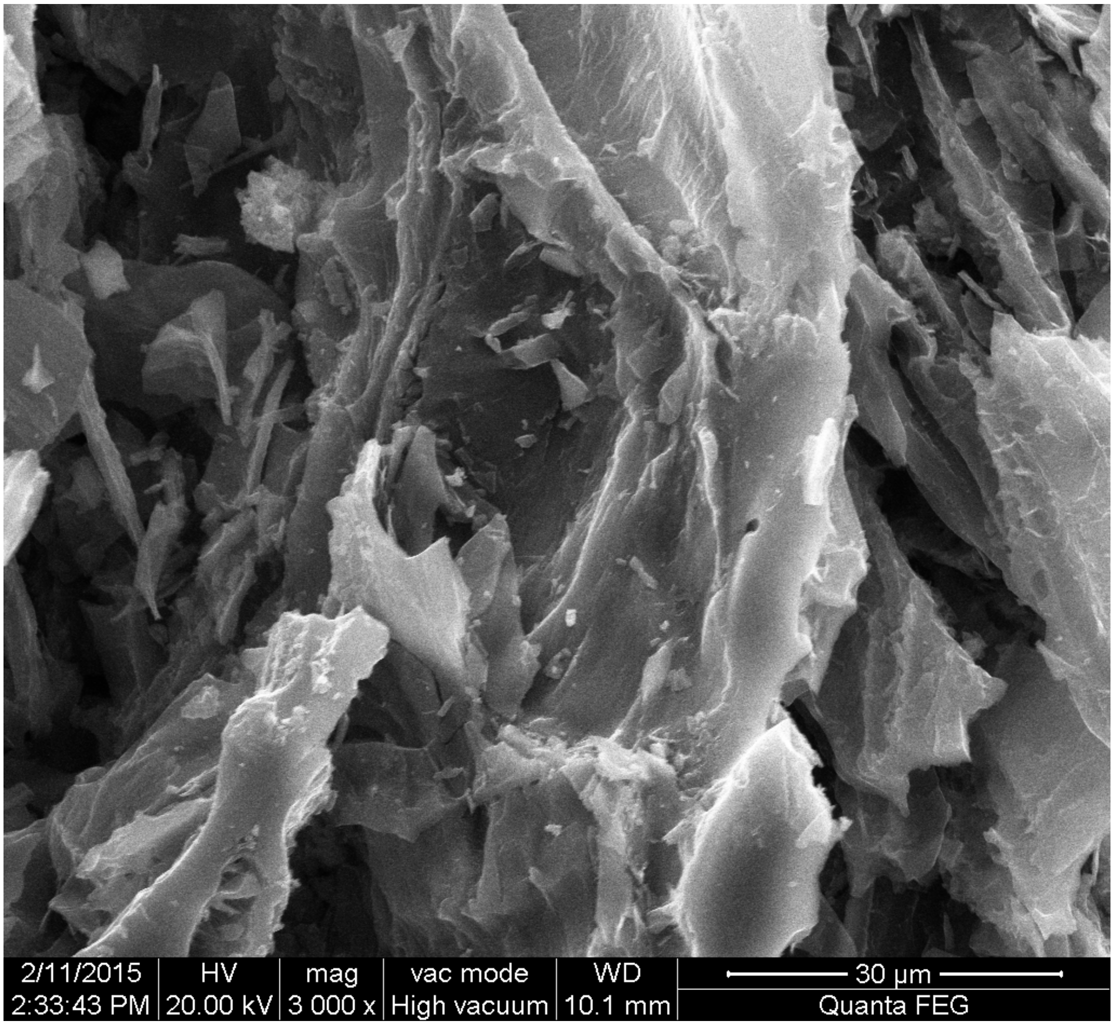


Figure 8 SEM analysis of CNSC after Dye treatment

EDS Chart
Spectrum: training
El AN Series unn. C norm. C Atom. C Error (1 Sigma)
[wt.%] [wt.%] [at.%] [wt.%]

C 6 K-series 100.00 100.00 100.00 14.98

Total: 100.00 100.00 100.00

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

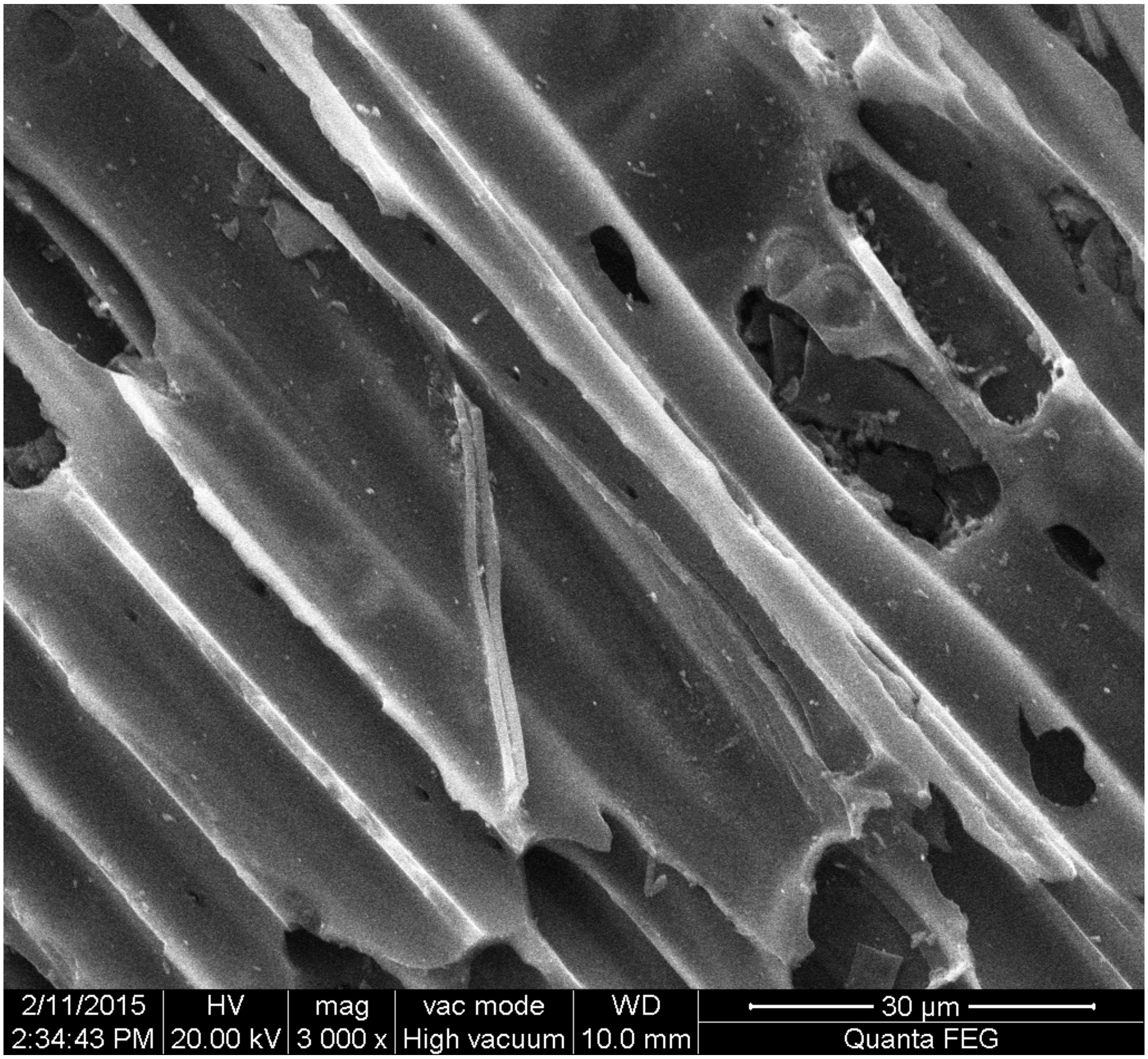


Figure 9 SEM analysis of CAC before Dye treatment

Table EDS Chart
Spectrum: training
El AN Series unn. C norm. C Atom. C Error (1 Sigma)
[wt.%] [wt.%] [at.%] [wt.%]

C 6 K-series 100.00 100.00 100.00 14.88

Total: 100.00 100.00 100.00

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

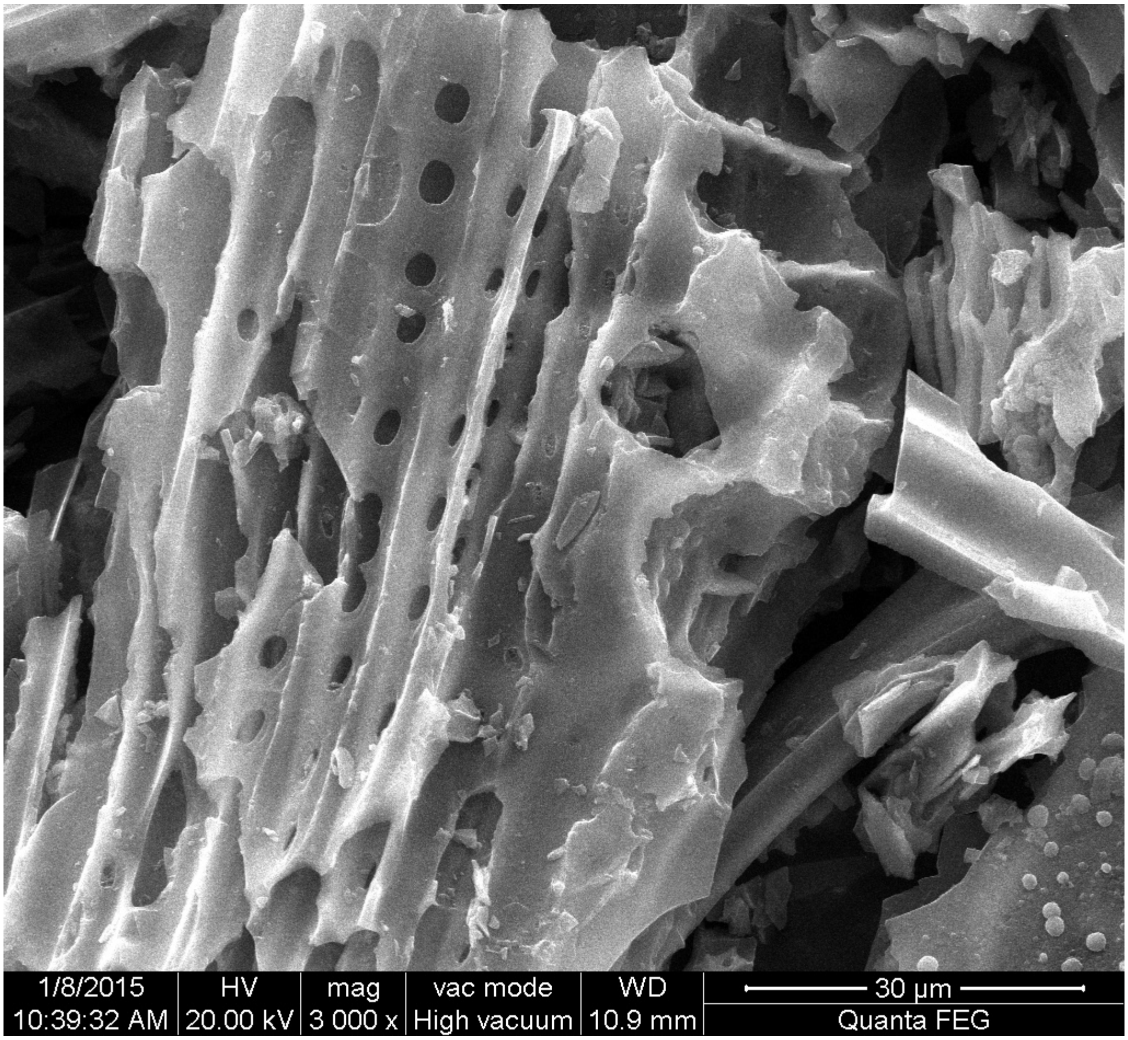


Figure 10 is the SEM analysis of CAC after treated with the dye.

EDS CHART

Spectrum: training

El AN Series unn. C norm. C Atom. C Error (1 Sigma)

[wt.%] [wt.%] [at.%] [wt.%]

C 6 K-series 100.00 100.00 100.00 13.89

Total: 100.00 100.00 100.00

DOI: 10.48175/IJARSCT-28706

XRD Analysis

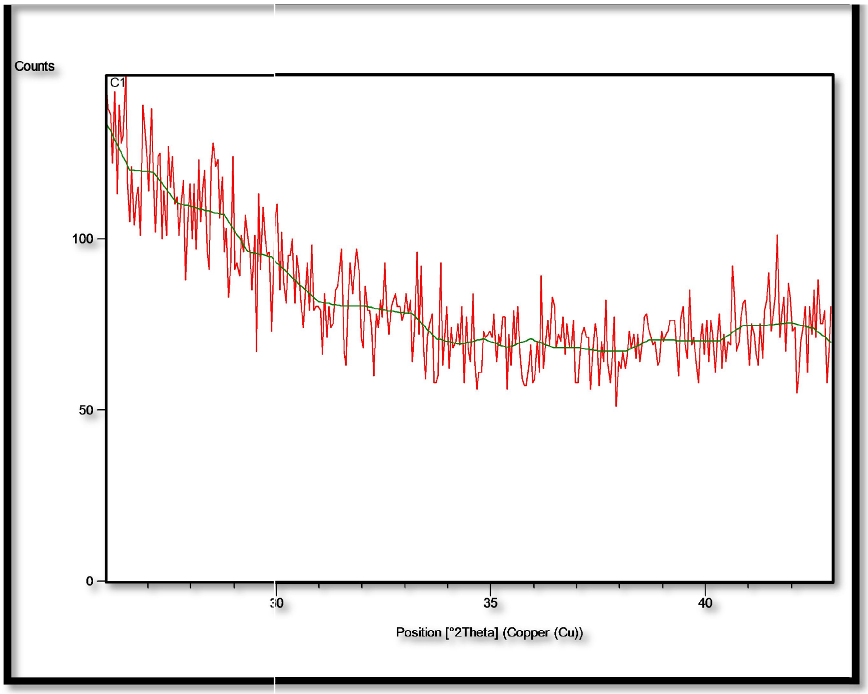


Figure 11 represents the XRD analysis of CNSC

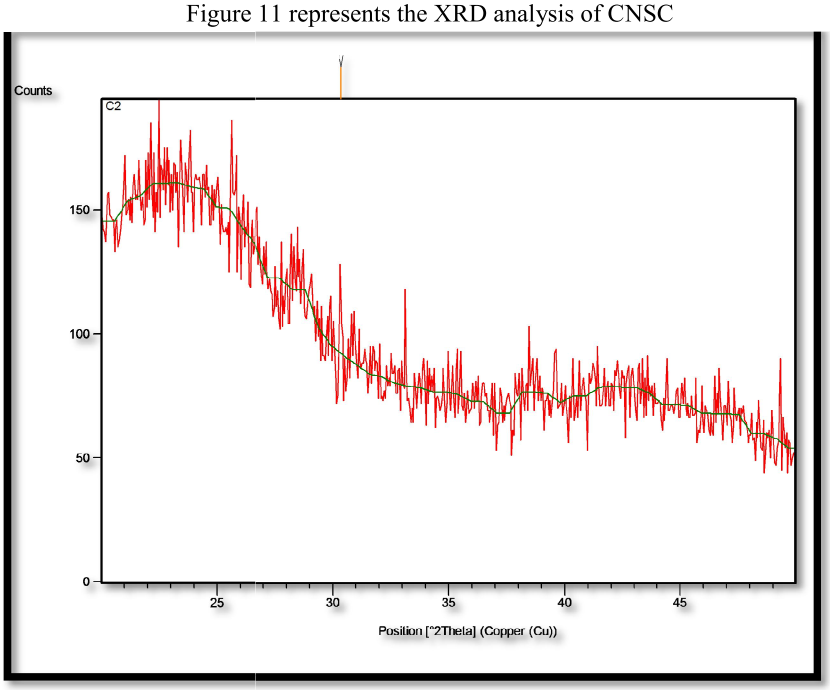


Figure 12 represents the XRD analysis of CNSC -Treated

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

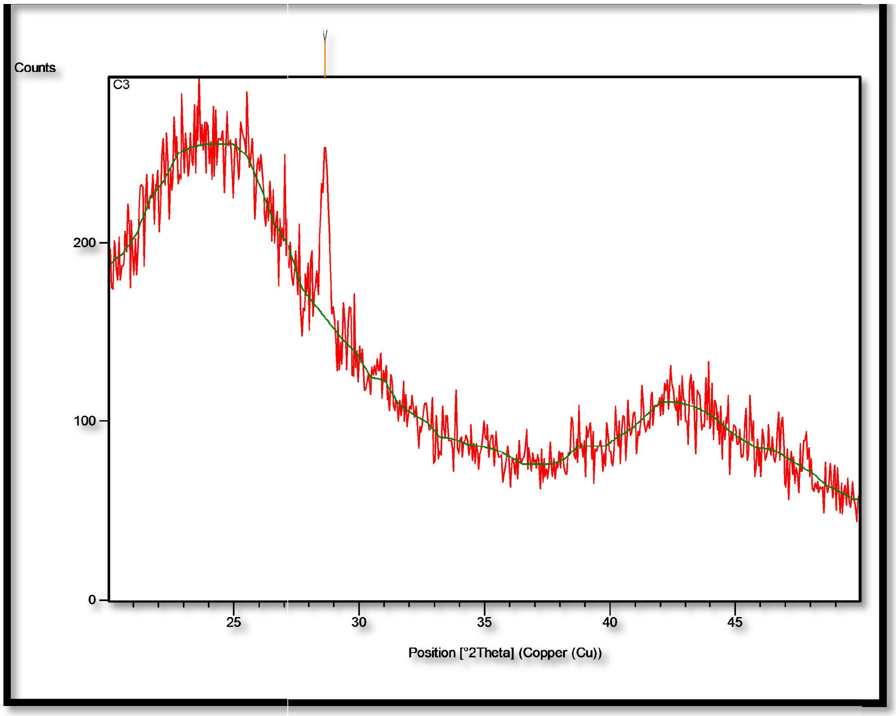


Figure 13 represents the XRD analysis of CAC

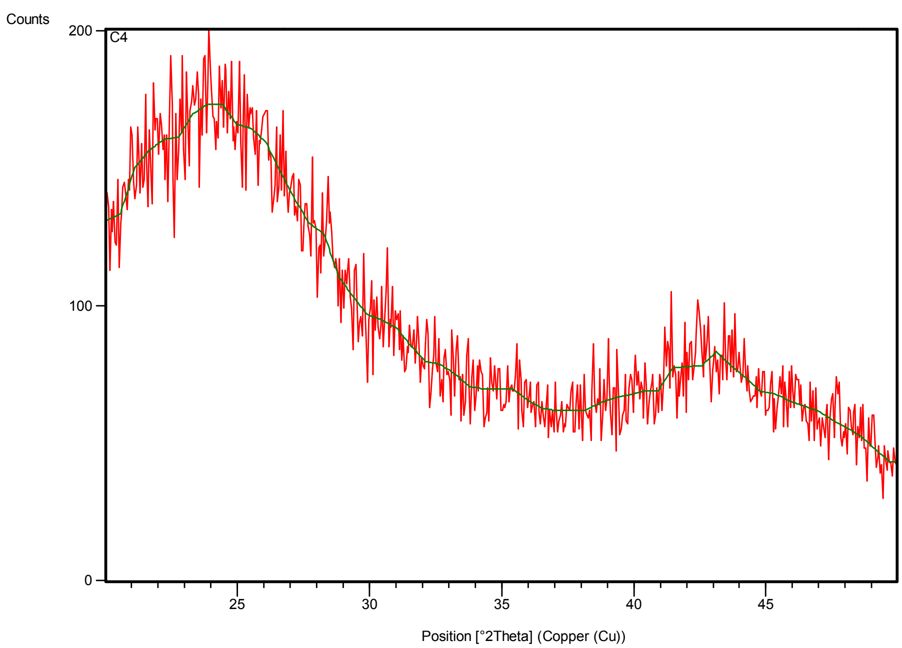


Figure 14 represents the XRD analysis of CAC –Treated

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

FTIR ANALYSIS

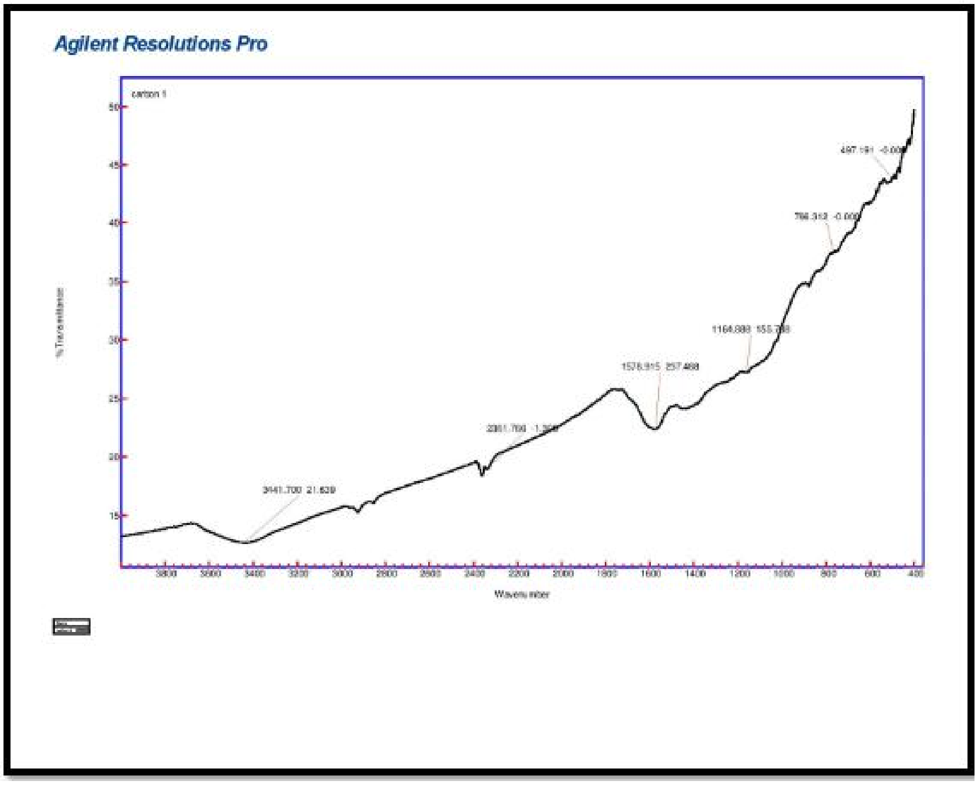


Figure 15 FTIR analysis of CNSC before treatment

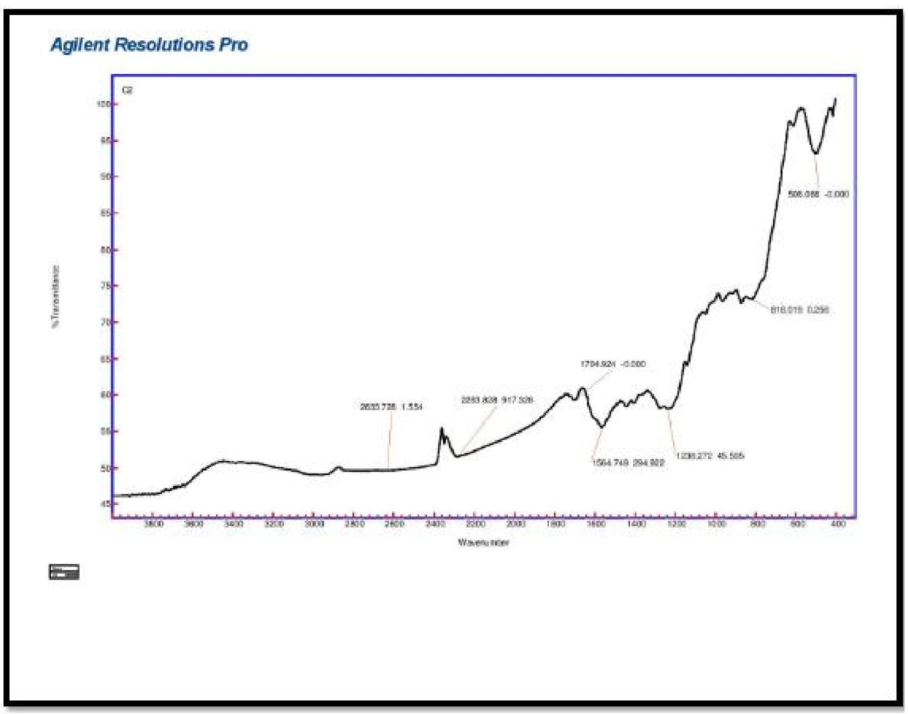


Figure 16 represents CNSC after dye treatment

DOI: 10.48175/IJARSCT-28706

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

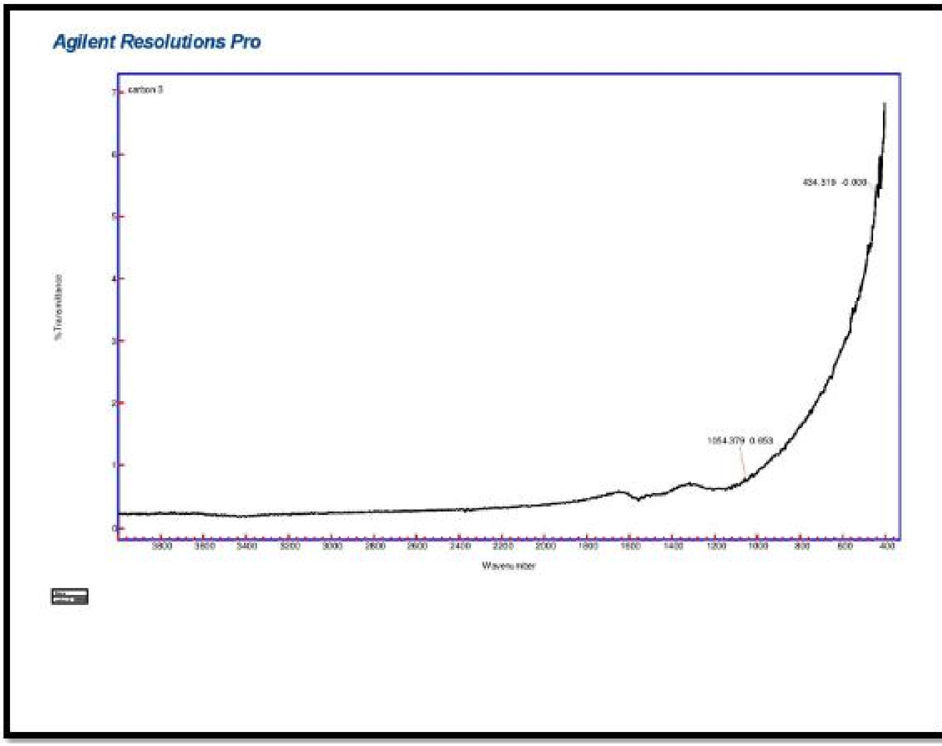


Figure 17 represents FTIR for CAC before treatment

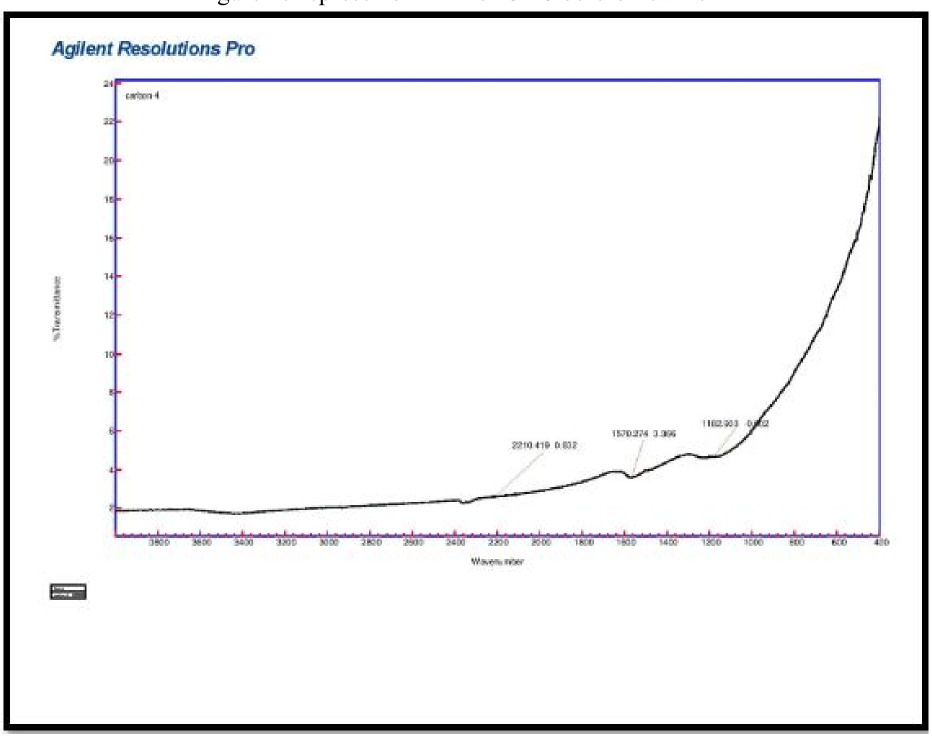


Figure 18 represents FTIR for CAC treated with the dye

DOI: 10.48175/IJARSCT-28706