Prediction of Shear Strength of Soft Soil using Python

Adwaith A.*
Dept. of Civil Engineering
Bishop Jerome Institute
Kollam, Kerala
aadwaith296@gmail.com

Arjun A.
Dept. of Civil Engineering
Bishop Jerome Institute
Kollam, Kerala
arjunsooraj14@gmail.com

Bijoy Fernandez
Dept. of Civil Engineering
Bishop Jerome Institute
Kollam, Kerala
bijoyfernandez@gmail.com

Josna J.
Dept. of Civil Engineering
Bishop Jerome Institute
Kollam, Kerala
josnajhnsn@gmail.com

Shobana
Asssistant Professor
Dept. of Civil Engineering
Bishop Jerome Institute
Kollam, Kerala
dhashanamerry4@gmail.com

* Corresponding author
doi: https://doi.org/10.21467/proceedings.7.8.4

ABSTRACT
Shear strength is a crucial parameter in determining the stability of soil structures. Accurate prediction of this parameter can significantly reduce the time and cost involved in conventional testing methods. This study presents a machine learning-based model developed in Python to predict the shear strength of soft soils using geotechnical properties such as moisture content, liquid limit, plastic limit, and specific gravity. Data was collected from different districts and processed to train the model. The results demonstrate the potential of Python-based algorithms in forecasting shear strength with a reasonable degree of accuracy.

I. Introduction

Soil shear strength is a crucial parameter in geotechnical engineering as it governs the stability of structures built on or within the soil. Accurate prediction of shear strength is essential for the safe design of foundations, slopes, and retaining walls. However, conventional testing methods such as the tri-axial and unconfined compression tests are labor-intensive and not always feasible. Machine learning offers a viable alternative for estimating soil strength parameters quickly and accurately using available soil properties. From the model implemented an application named ShearMate is also developed and is available in an open online platform called Github

II. Literature Review

Several studies have explored the application of ML and artificial neural networks (ANNs) in predicting the geotechnical properties of soils. Bhuvaneswari et al. [1] demonstrated the effectiveness of fly ash in stabilizing expansive soils, highlighting the importance of additives in improving soil behavior. Rahman et al. [2] employed ANNs to estimate the shear strength of soils and found that models performed accurately when trained with proper geotechnical parameters. Similarly, Santosh and Sarma [3] extended ANN applications to unsaturated soils, achieving reliable predictions for shear strength parameters. In another study, Jakka [4] investigated the mechanical behavior of soft clays, emphasizing their low shear strength and high compressibility. Bera and Ghosh [5] applied ANN models to predict shear strength in silty clay, reinforcing the role of data-driven techniques in geotechnical engineering. Recent advancements have focused on improving prediction accuracy through ensemble learning and advanced ML algorithms. Nguyen et al. [6] applied Random Forest (RF) and Support Vector Regression (SVR) to predict UCS with high accuracy, underlining the importance of proper data preprocessing. Saha and Suman [7] demonstrated the use of deep learning (DL) models for soil classification and shear strength estimation from limited input data. More recently, Bhattacharya et al. [8] developed ensemble models for predicting soil properties and confirmed that ML techniques consistently outperformed conventional regression methods in terms of both speed and accuracy. These findings underline the growing relevance of ML and AI-based approaches in geotechnical engineering. By leveraging soil index properties, such models have the potential to significantly reduce dependence on time-intensive laboratory testing.

Proceedings DOI: 10.21467/proceedings.7.8; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-989164-8-8

III. Problem Identification and Methodology

A. *Identified problems vs. Adopted methodology*

Soft soils, common in regions like Kuttanad and Kollam, exhibit low shear strength and high compressibility, posing significant challenges for construction. Traditional strength determination methods like UCS testing are time-consuming and require extensive lab work. To address this, the study adopts an AI-based approach using Python to predict shear strength based on soil index properties. Fifty samples were collected and tested for parameters such as specific gravity, liquid limit, plastic limit, and initial moisture content. These values were used to train machine learning models, aiming to develop a reliable, quick, and cost-effective prediction tool that reduces dependency on destructive testing methods.

B. *Objectives of the Methodology*

The primary objective of this study is to develop a machine learning model that can accurately predict the shear strength of soft soils using basic index properties. This approach aims to minimize the need for time-consuming and labor-intensive laboratory tests like the unconfined compressive strength (UCS) test. By analyzing parameters such as specific gravity, liquid limit, plastic limit, and moisture content, the study seeks to establish strong correlations that can be leveraged for predictive modeling. Ultimately, the methodology is intended to provide a faster, more economical, and non-destructive alternative for geotechnical strength assessment, particularly useful in soft soil-dominated regions.

IV. Data Collection and Validation

A. *Data Parameters Collected*

Table 1 Final dataset passed onto the prediction model

Sample	LL(%)	PL(%)	IMC (%)	SG	UCT (kg/cm^2)	Loc
1	40.184	20.143	48.648	2.74	0.11005	KND
2	62.044	35.427	43.018	2.67	0.11185	KND
3	72.759	41.365	41.594	2.64	0.11075	KND
4	45.759	18.305	38.946	2.56	0.11183	KND
5	54.016	34.055	91.232	2.68	0.10793	KND
6	60.374	28.704	52.261	2.67	0.10635	KND
7	74.979	31.479	37.292	2.75	0.1088	KND
8	72.232	36.058	56.304	2.66	0.10368	MKD
9	65.462	40.885	51.59	2.56	0.10523	MKD
10	59.929	39.611	57.215	2.57	0.10383	MKD
11	56.267	29.425	52.468	2.58	0.1063	MKD
12	78.121	33.8	65.418	2.64	0.10345	MKD
13	60.902	40.119	102.829	2.53	0.10055	MKD
14	48.765	21.151	66.321	2.65	0.10303	MKD
15	35.754	22.833	43.751	2.83	0.11495	CGP
16	50.137	25.733	34.496	2.82	0.11878	CGP
17	44.275	19.8	49.492	2.69	0.10818	CGP
18	39.468	24.063	40.045	2.83	0.1167	CGP
19	62.968	24.667	27.127	2.83	0.11725	CGP
20	50.286	21.681	34.386	2.79	0.1152	CGP
21	30.198	34.664	53.059	2.87	0.116	CGP
22	37.425	13.738	21.05	2.79	0.11938	MYL
23	59.451	21.899	43.621	2.82	0.10965	MYL
24	43.68	13.395	32.34	2.74	0.11363	MYL
25	59.586	26.006	46.212	2.81	0.10998	MYL
26	30.819	35.779	39.139	2.65	0.11655	MYL
27	51.029	21.872	45.009	2.83	0.11473	MYL
28	46.69	23.619	34.983	2.74	0.11455	MYL

Here the data of 7 samples each from 4 different locations is being noted and the rest of the data is shown in Appendix-1

B. Data Collection and Validation

Soil samples were collected from three districts in Kerala—Kollam, Kuttanad, and Kannur—known for soft clayey soils. A total of 120 samples were tested in the laboratory for key index properties including specific gravity, liquid limit, plastic limit, and initial moisture content. The unconfined compressive strength (UCS) of each sample was also determined for validation. The dataset was carefully reviewed to eliminate outliers and ensure consistency. These validated results were then used to train and test machine learning models in Python, with a focus on predicting shear strength based on the collected soil parameters. The approach ensures data reliability for effective model training.

C. Laboratory Testing and Data Validation

Laboratory testing was conducted on 50 soft soil samples collected from Kollam, Kuttanad, and Kannur. Each sample underwent standard geotechnical tests to determine specific gravity, liquid limit, plastic limit, and initial moisture content. Unconfined compressive strength (UCS) was also measured to serve as the target output for the AI model. Data validation involved checking for consistency, removing anomalies, and ensuring accuracy across all parameters. The verified dataset was then normalized and prepared for model training and testing, ensuring the reliability of the results used in developing an accurate and efficient shear strength prediction model.

D. Equations

$$pi = ll - pl \qquad (1)$$

$$imc = {}^{\mathrm{Ww}}/_{\mathrm{Wd}} \times 100 \qquad (2)$$

V. MODEL IMPLEMENTATION

A. Data pre-processing and correlation analysis

Data preprocessing was carried out to prepare the soil test results for machine learning model development. This included organizing the collected data into a structured format and handling missing or inconsistent values. Outliers were identified and removed to improve model accuracy. The input features—specific gravity, liquid limit, plastic limit, and moisture content—were normalized to ensure uniformity and better algorithm performance. The dataset was then split into training and testing sets, maintaining a suitable ratio to validate the model effectively. These steps ensured that the data used for model training was clean, consistent, and suitable for predictive analysis.

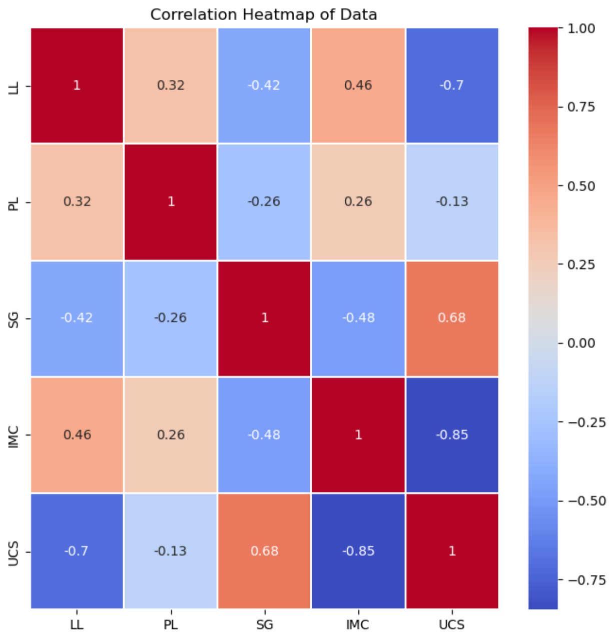

Fig. 1 heatmap representing accuracy of the model

B. Model Training and Performance Evaluation

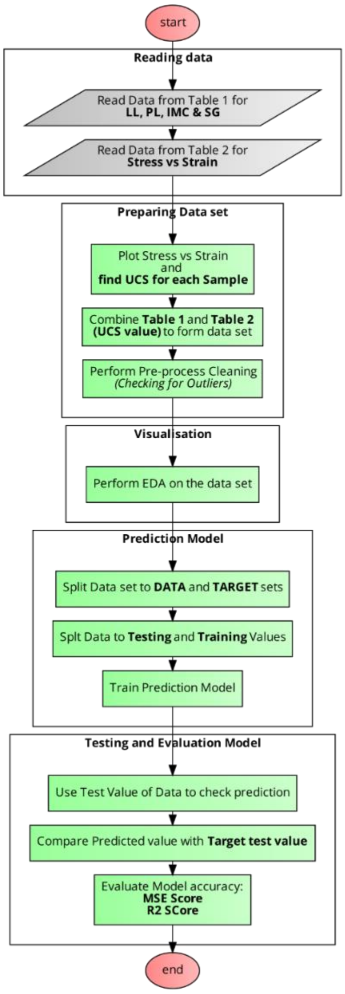

fig .2 ML model process

The preprocessed dataset was used to train machine learning models in Python to predict the shear strength of soft soils. Various regression algorithms were tested to identify the most suitable model based on performance metrics. The data was split into training and testing sets, and models were trained using input features like liquid limit, plastic limit, specific gravity, and moisture content. Performance was evaluated using statistical indicators such as R^2 score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The best-performing model demonstrated high accuracy, validating the effectiveness of AI in soil strength prediction. The dataset, consisting of soil parameters such as liquid limit, plastic limit, specific gravity, and moisture content, was split into 80% training and 20% testing. Hyperparameters like the number of trees (100), maximum depth (10), and minimum samples per leaf (1) were optimized using GridSearchCV. The model's performance was evaluated using R^2 (0.873) and MSE (0.000105), confirming its suitability for accurate and efficient shear strength prediction in geotechnical applications. The Random Forest model effectively predicts UCS values based on key soil properties, showing good alignment with expected trends. Minor discrepancies suggest room for improvement through additional features or hyperparameter tuning. Overall, the approach demonstrates the practical value of ML in geotechnical analysis and decision-making.

Proceedings DOI: 10.21467/proceedings.7.8
ISBN: 978-81-989164-8-8

Series: AIJR Proceedings
ISSN: 2582-3922

C. Model Development and Evaluation

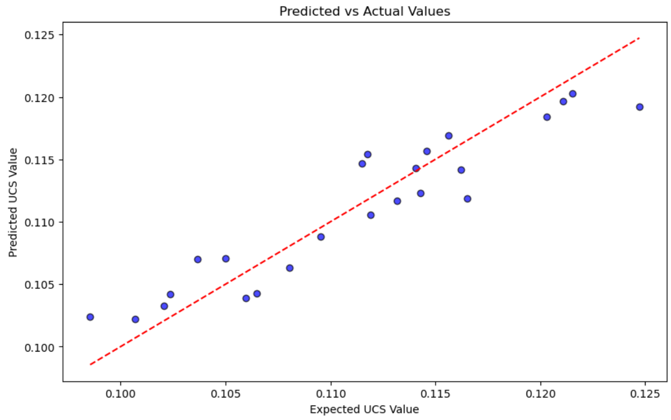

Fig. 3 predicted vs actual ucs

Several regression models including Linear Regression, Support Vector Regression, and Random Forest were trained. Model performance was evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2). The Random Forest model outperformed others with the lowest error values and highest R^2 score. Hyperparameter tuning was performed using GridSearchCV to optimize the model.

VI. APPLICATION INTERFACE – SHEARMATE

An intuitive Python-based interface named ShearMate was developed to make the prediction model user-friendly for engineers and field professionals. The application allows users to input basic soil parameters—such as liquid limit, plastic limit, specific gravity, and moisture content—and instantly receive predicted shear strength values. This tool simplifies the analysis process, eliminates the need for complex testing, and enhances decision-making in geotechnical design and planning. ShearMate demonstrates the practical utility of integrating machine learning with geotechnical engineering applications

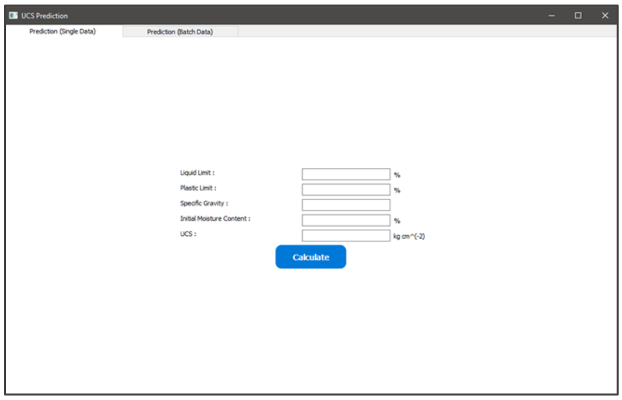

Fig. 4 Single Sample Input Page

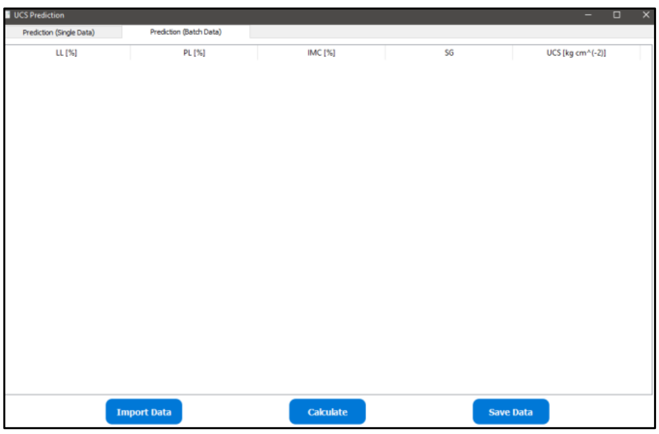

fig .5 Bulk Data Entry Page

A. Merits, Demerits, and Limitations

ShearMate offers a fast, cost-effective, and user-friendly solution for predicting shear strength without extensive lab testing. It enables quick decision-making and is ideal for preliminary site assessments. However, its accuracy depends on the quality and range of input data. The current model is limited to the types of soils and parameters used in training, and may not generalize well to all soil conditions. Further enhancements with larger and more diverse datasets can improve its reliability and applicability.

B. Availability of ShearMate

ShearMate is developed as a standalone Python application and is available upon request for academic and research purposes. The tool can be shared as an executable file or Python script, enabling easy access and customization for users interested in integrating machine learning into geotechnical soil strength prediction the application is now available in an online platform called github and is easy to install and use.

VII. RESULTS

The developed model demonstrated high accuracy in predicting shear strength, with strong correlation between predicted and actual UCS values. Performance metrics such as R^2 score and low error values confirmed the model's reliability, indicating its potential as an effective tool for preliminary geotechnical assessments using basic soil parameters.

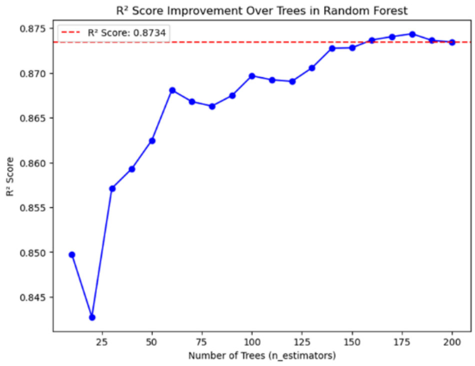

Fig. 6 Predicted vs Actual Values of UCS

A. Model Assessment and Comparison with Conventional Methods

The AI-based model was assessed for accuracy and reliability by comparing its predictions with conventional UCS testing results. The model showed strong alignment with lab-tested values, significantly reducing time and cost involved in soil strength estimation. While traditional methods require extensive sampling and equipment, the ML model provides fast and fairly accurate predictions using basic soil parameters. Although not a complete replacement, it serves as a powerful preliminary assessment tool in geotechnical engineering, especially in resource-limited scenarios.

B. Conclusion of results

The study successfully demonstrated the potential of machine learning in predicting the shear strength of soft soils using fundamental geotechnical parameters. The developed model yielded reliable and accurate results, aligning closely with conventional UCS test outcomes. This approach offers a faster and cost-effective alternative for preliminary soil strength evaluation, making it a valuable tool for geotechnical engineers in early-stage project planning.

VIII. CONCLUSION

This project highlights the successful use of machine learning for predicting the shear strength of soft soils using fundamental geotechnical parameters like moisture content, liquid limit, plastic limit, and specific gravity. The developed Python-based model, deployed through the *ShearMate* interface, demonstrated high accuracy and efficiency in estimating UCS values, closely matching laboratory results. It significantly reduces time, effort, and cost involved in conventional testing. While the model is intended for preliminary assessments and not as a full replacement for lab procedures, it serves as a practical and accessible decision-support tool for engineers. With further data and refinement, the system holds potential for broader geotechnical applications and enhanced prediction capabilities. However, the model is limited to the specific dataset used and may not generalize to all soil types. Future work will focus on expanding the dataset, improving model robustness, and exploring deep learning approaches for better performance across varied soil conditions.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the faculty and staff of the department of civil engineering, Bishop Jerome Institute, for their valuable support and guidance throughout the course of this project. Special thanks to our project guide, sr. Sobhana, for her continuous encouragement, insightful feedback, and expertise that significantly contributed to the successful completion of this study.
The authors also extend their appreciation to the laboratory technicians and supporting staff for their assistance in conducting the experimental investigations.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest in this study.

REFERENCES

[1] S. Bhuvaneshwari, R. G. Robinson, and S. R. Gandhi, "Stabilization of expansive soils using fly ash," Fly Ash India, vol. 8, no. 5, pp. 1–10, 2005. [Online]. Available: https://www.researchgate.net/publication/282293593
[2] M. S. Rahman, M. T. Alamgir, and A. S. M. A. Haseeb, "Prediction of shear strength of soil using artificial neural network," Journal of Civil Engineering (IEB), vol. 37, no. 2, pp. 133–143, 2009.
[3] K. C. Santosh and A. L. Sarma, "Artificial neural network for prediction of shear strength parameters of unsaturated soils," Geotechnical and Geological Engineering, vol. 31, no. 2, pp. 651–658, Apr. 2013. [Online]. Available: https://doi.org/10.1007/s10706-012-9593-5
[4] M. B. Jakka, "Shear strength behavior of soft clays," Indian Geotechnical Journal, vol. 41, no. 3, pp. 199–206, Jul. 2011. [Online]. Available: https://doi.org/10.1007/s40098-011-0016-y
[5] S. Bera and S. Ghosh, "Shear strength prediction of silty clay soil using ANN model," in Proc. Indian Geotechnical Conf., Dec. 2015, pp. 113–116.
[6] M. D. Nguyen, H. T. Bui, and D. T. Tran, "Machine learning techniques in predicting unconfined compressive strength of soils," J. Rock Mech. Geotech. Eng., vol. 12, no. 5, pp. 1084–1095, 2020. [Online]. Available: https://doi.org/10.1016/j.jrmge.2020.01.007
[7] S. Saha and S. Suman, "Deep learning model for soil classification and strength estimation," Int. J. Geotech. Eng., vol. 15, no. 4, pp. 367–375, 2021. [Online]. Available: https://doi.org/10.1080/19386362.2019.1669554
[8] B. Bhattacharya, R. Singh, and A. Sharma, "Ensemble learning-based prediction of soil properties for geotechnical applications," Engineering Geology, vol. 307, p. 106589, 2022. [Online]. Available: https://doi.org/10.1016/j.enggeo.2022.106589

Appendix 1

Table 1 Final dataset passed onto the prediction model

Samples	LL	PL	IMC (%)	SG	UCT (kg/cm^2)	Location
1	40.184	20.143	48.648	2.74	0.11005	KND
2	62.044	35.427	43.018	2.67	0.11185	KND
3	72.759	41.365	41.594	2.64	0.11075	KND
4	45.759	18.305	38.946	2.56	0.111825	KND
5	54.016	34.055	91.232	2.68	0.107925	KND
6	60.374	28.704	52.261	2.67	0.10635	KND
7	74.979	31.479	37.292	2.75	0.1088	KND
8	39.192	19.134	46.477	2.6	0.109425	KND
9	54.943	18.075	46.032	2.66	0.107275	KND
10	50.405	33.672	40.452	2.68	0.1148	KND
11	69.633	37.583	48.065	2.68	0.1081	KND
12	54.039	36.55	44.963	2.66	0.11405	KND
13	39.439	30.595	45.254	2.71	0.11655	KND
14	60.377	19.497	57.171	2.59	0.103975	KND
15	65.831	34.201	41.698	2.7	0.113025	KND
16	35.115	23.047	54.256	2.67	0.11215	KND
17	69.358	39.777	52.119	2.73	0.109525	KND

18	42.441	37.957	52.043	2.72	0.111925	KND
19	57.083	34.735	54.89	2.61	0.10725	KND
20	51.632	40.721	51.451	2.61	0.110375	KND
21	63.786	17.786	61.378	2.75	0.10655	KND
22	49.846	35.405	55.267	2.66	0.10955	KND
23	125.055	25.256	51.409	2.73	0.109075	KND
24	64.22	39.652	59.893	2.58	0.105975	KND
25	60.909	23.667	53.184	2.68	0.105	KND
26	66.85	35.121	43.734	2.71	0.113175	KND
27	65.161	69.09	41.213	2.74	0.112975	KND
28	36.438	18.041	39.57	2.62	0.11175	KND
29	50.789	26.941	61.984	2.69	0.10805	KND
30	72.232	36.058	56.304	2.66	0.103675	MKD
31	65.462	40.885	51.59	2.56	0.105225	MKD
32	59.929	39.611	57.215	2.57	0.103825	MKD
33	56.267	29.425	52.468	2.58	0.1063	MKD
34	78.121	33.8	65.418	2.64	0.10345	MKD
35	60.902	40.119	102.829	2.53	0.10055	MKD
36	48.765	21.151	66.321	2.65	0.103025	MKD
37	58.359	40.202	69.231	2.59	0.10315	MKD
38	76.775	24.447	40.826	2.58	0.1063	MKD
39	47.044	32.867	65.194	2.7	0.1088	MKD
40	48.558	40.207	62.553	2.61	0.107475	MKD
41	40.521	26.346	46.861	2.54	0.110725	MKD
42	40.568	26.145	68.69	3.82	0.1076	MKD
43	73.176	30.069	44.445	2.53	0.1067	MKD
44	69.285	42.366	60.386	2.67	0.1079	MKD
45	95.341	20.011	46.747	2.67	0.107675	MKD
46	67.932	42.535	52.941	2.59	0.1057	MKD
47	55.808	23.845	61.575	2.68	0.1039	MKD
48	78.116	44.752	56.9	2.62	0.1043	MKD
49	78.682	43.871	52.825	2.61	0.108175	MKD
50	55.712	32.367	49.79	2.53	0.106325	MKD
51	74.169	32.818	65.19	2.58	0.09945	MKD
52	70.924	23.039	44.314	2.71	0.110525	MKD
53	52.4	35.886	52.903	2.6	0.108875	MKD
54	49.282	22.268	67.113	2.63	0.10235	MKD
55	58.329	20.815	65.375	2.63	0.10205	MKD
56	71.254	21.443	52.028	2.64	0.10365	MKD
57	78.407	31.109	62.893	2.53	0.1007	MKD
58	70.543	26.889	64.695	2.59	0.09855	MKD
59	74.208	35.551	122.372	2.66	0.1016	MKD
60	64.123	44.811	56.335	2.56	0.106475	MKD
61	35.754	22.833	43.751	2.83	0.11495	CGP
62	50.137	25.733	34.496	2.82	0.118775	CGP
63	44.275	19.8	49.492	2.69	0.108175	CGP
64	39.468	24.063	40.045	2.83	0.1167	CGP
65	62.968	24.667	27.127	2.83	0.11725	CGP
66	50.286	21.681	34.386	2.79	0.1152	CGP
67	30.198	34.664	53.059	2.87	0.116	CGP
68	47.404	12.402	50.52	2.74	0.109475	CGP
69	56.229	31.554	54.457	2.86	0.113125	CGP
70	53.34	15.11	43.84	2.85	0.1138	CGP
71	31.884	31.939	32.425	2.67	0.11765	CGP
72	63.897	27.398	29.884	2.76	0.114775	CGP
73	42.953	33.082	30.607	2.87	0.12085	CGP
74	31.361	14.438	25.594	2.69	0.1185	CGP
75	32.208	25.975	36.72	2.78	0.1189	CGP

Proceedings DOI: 10.21467/proceedings.7.8
ISBN: 978-81-989164-8-8

Series: AIJR Proceedings
ISSN: 2582-3922

76	28.929	24.527	30.743	2.87	0.122275	CGP
77	57.141	29.655	37.35	2.78	0.113325	CGP
78	29.881	25.357	99.515	2.71	0.110675	CGP
79	30.005	48.255	37.916	2.79	0.117	CGP
80	59.019	22.74	51.34	2.71	0.106225	CGP
81	38.784	19.721	43.165	2.77	0.114325	CGP
82	26.234	24.057	32.414	2.68	0.118825	CGP
83	27.076	14.962	35.676	2.79	0.11665	CGP
84	30.687	36.038	27.518	2.77	0.124725	CGP
85	33.536	29.86	35.004	3.8	0.11765	CGP
86	51.606	35.812	45.326	2.8	0.116225	CGP
87	31.293	17.668	27.163	2.78	0.1203	CGP
88	29.755	33.984	31.162	2.81	0.12155	CGP
89	55.589	32.22	36.443	2.7	0.1115	CGP
90	51.842	28.929	54.461	2.79	0.111925	CGP
91	37.425	13.738	21.05	2.79	0.119375	MYL
92	59.451	21.899	43.621	2.82	0.10965	MYL
93	43.68	13.395	32.34	2.74	0.113625	MYL
94	59.586	26.006	46.212	2.81	0.109975	MYL
95	30.819	35.779	39.139	2.65	0.11655	MYL
96	51.029	21.872	45.009	2.83	0.114725	MYL
97	46.69	23.619	34.983	2.74	0.11455	MYL
98	51.576	18.384	36.051	2.76	0.11355	MYL
99	55.999	27.927	24.332	2.7	0.12015	MYL
100	48.6	43.761	43.52	2.8	0.115825	MYL
101	41.922	22.305	32.582	2.67	0.1166	MYL
102	51.713	21.326	35.002	2.82	0.11585	MYL
103	35.082	35.124	43.553	2.82	0.11995	MYL
104	108.661	22.582	37.965	2.69	0.110325	MYL
105	25.372	25.338	26.033	2.81	0.123125	MYL
106	54.288	20.083	37.707	2.81	0.113225	MYL
107	56.453	32.669	32.229	2.78	0.11595	MYL
108	54.69	22.936	34.086	3.83	0.1147	MYL
109	42.509	25.84	20.849	2.7	0.120925	MYL
110	39.535	19.334	26.301	2.7	0.11505	MYL
111	61.85	37.691	28.431	2.71	0.116525	MYL
112	31.167	20.287	41.389	2.83	0.11805	MYL
113	43.875	32.269	20.345	2.67	0.117225	MYL
114	48.827	23.687	32.263	2.82	0.115625	MYL
115	50.035	18.155	47.721	3.87	0.109125	MYL
116	30.161	32.993	23.356	2.75	0.1211	MYL
117	48.16	21.34	34.493	2.81	0.114575	MYL
118	54.735	34.63	43.596	2.83	0.11405	MYL
119	24.001	35.881	48.276	2.72	0.116525	MYL
120	58.461	34.496	42.403	2.74	0.1143	MYL

CGP - Chungapalam; MYL - Mylom; MKD – Mukkada bridge; KND - Kaniyamthodu

Proceedings DOI: 10.21467/proceedings.7.8
ISBN: 978-81-989164-8-8

Series: AIJR Proceedings
ISSN: 2582-3922