Predicting the Quality of Life Index Value for determining the Requirement of Support Needs for Intellectually Disabled Individual using Machine Learning Methods

Gaurav Kumar Yadav 1,2*, Benigno Moreno Vidales 3*, Hatem A. Rashwan 2*, Joan Oliver 3*, Mohamed Abdel-Nasser 4*, Domenec Puig 2*, G. C. Nandi 1*

1* *Department of Information Technology, Indian Institute of Information Technology Allahabad, Prayagraj, 211012, India*

2* *Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Tarragona, 43002, Spain*

3* *Instituto de Robótica para la Dependencia, Sitges, 08870, Barcelona, Spain*

4* *Electronics and Communication Engineering Section, Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan, 81528, Egypt*

***Abstract*—Recently, the number of cases of intellectually disabled people has been increasing. People with intellectual disability (ID) in their older age face many challenges that ordinary people do not. Each elder individual needs a different kind to support to improve their lifestyle. The concept of quality of life (QoL) in intellectual disability provides a trait to improve the individual QoL by providing support. There is a set of 69 questionnaires related to every aspect of the individual life of an ID person. Based on the questionnaire response, professionals calculate the QoL index value, which decides Who needs support and who doesn't. So far, professionals have used the GENCAT tool to calculate the QoL index value by initially converting the 69 items' responses into the eight dimensions value and, after that, from the eight dimensions value to the QoL index. This process is lengthy and tedious and needs an expert to calculate the index value. This work proposes using a trained state-of-the-art machine learning (ML) based model to directly predict the QoL index value by inputting the response of the 69 items. This paper proposes different ML-based models using the original and augmented NewtonOne dataset. Based on various evaluation metrics such as mean absolute error, root mean squared error and R^2 score value, we choose the Ridge regression algorithm to predict the QoL index value. The test case value of various evaluation metrics is $1.6676, 1.9643$, and 0.9745 for the MAE, RMSE, and R^2 score, respectively, for the ridge regression model.**

***Index Terms*—Quality of life, Intellectual disability, Ensemble learning, Machine learning, Regression technique.**

I. Introduction

In recent decades the population of intellectually disabled (ID) people has increased worldwide. ID people cover around 1 to 3 $\%$ of the global population [1]. This increasing figure motivates researchers and other professionals to work in this area to improve the lifestyle of ID people. ID is a severe, widespread neurodevelopmental condition that has lifelong effects. It appears early stage of life till 18 years of life [2]. ID people show inefficiency in their intellectual and adaptive behaviour. ID is measured by using an intelligence quotient (IQ) level score. It could show itself in mild to profound ways. If a kid has an IQ score of less than 70, considered an ID kid [3]. ID differentiate based on the IQ score, such as if a kid's IQ is less than 70 and higher than 50, categorised as mild ID. Around 85% of the total ID population belongs to this category. These children can finish their education, develop independence, get training, and possibly pursue employment. A moderate IQ falls between 35 and 50. Despite having the potential to function independently, they frequently need care and attention. A total of 10% of the ID population belongs to this category. A severe intellectual disability exists in cases with IQ scores between 20 and 35. These people are less competent and struggle to understand numbers and reading. They require regular supervision. 4% of the total ID population belongs to this category. A profound ID instance is defined as having an IQ below 20. 1% of the total ID people population belongs to this category [4]. The cause of ID is genetically inherited in some kids, whereas in some cases, it appears due to the effect of medication consumed to cure the disease at an early age. However, the cause of ID is unclear in some cases [5].

To support ID people to improve their lifestyle, they need to analyse their quality of life (QoL). QoL is multidimensional, having etic (universal) and emic (cultural bound) components. It has objective and subjective aspects and affects by personal and environmental factors [6]. Improving QoL leads to improving the three (personal, social and judicial) areas of ID people's life. These three areas are broad categories into the eight dimensions of QoL. Improving the QoL of a person requires improving these eight dimensions [7]. These eight dimensions are interpersonal relation (IR), emotional well-being (EW), physical well-being (PW), material well-being (MW), personal development (PD), social inclusion (IS), self-determination (SD), and rights (RI). These eight dimensions define the quality of life. Improvement in these shows the improvement in QoL. To measure the values of these eight dimensions using the response of questionnaires, Verdugo el proposed a tool known as the GENCAT scale [8]. The GENCAT scale calculates the eight dimension values using specified rules and following some tables. After that, using the eight dimensions value, the GENCAT tool uses to calculate the QoL index value. The QoL index value defines the need for support. QoL index value shows if a person needs support or not [9]. The index value is essential in improving an ID person's QoL. Calculating the index value using the GENCAT tool is tedious and requires knowledge. Generally, a professional needs to calculate the index value using the GENCAT tool. However the GENCAT tool is efficient and accurate in calculating the index value, but it is an entirely statistical method and requires time and professionalism to calculate the index value.

Therefore, in this work, we propose to use an ML-based model to calculate the index value directly from questionnaire responses. Whereas earlier, professionals need to calculate the eight dimensions value from the response of the 69 questionnaires and then from eight dimensions value to the index value using the GENCAT tool. ML models learn the weights and biases parameters attached to the input features during training. After training, the trained ML-based models predict the QoL index value accurately. In this study, we trained ML-based models using the Newton One dataset. Various regression-based ML models are assessed and compared during training and testing. Some ML-based models' performances are close to each other.

- Unlike existing manual approaches for predicting QoL, we propose an accurate ML-based model for predicting QoL automatically in this paper.
- Providing comparative study between 15 ML-based models for QoL index prediction.

This paper is organised into five sections. Section two speaks about the background of the work. Whereas in section three, we show the methodology and experimental details of the paper. After that, section four depicts the results of the work. And finally, section five concludes our work and proposes future aspects related to this work.

II. Background

A. Quality of life and its dimensions

A person's QoL is a multifaceted phenomenon of fundamental domains influenced by personal and environmental factors. Although they may differ in relative worth and significance, these fundamental areas apply to everyone [10]. The present method of measuring QoL is distinguished by: (1) its multidimensional nature, which includes fundamental areas and metrics; (2) the use of many methodologies, including both subjective and objective metrics; (3) the application of a systems viewpoint that takes into account the various contexts that have an impact on people at the meso, micro, and macrosystem levels; (4) the greater participation of those at risk of being socially excluded in the planning and execution processes [11]. QoL has eight dimensions, emotional well-being (EW), interpersonal relation (IR), physical well-being(PW), social inclusion (SI), personal development (PD), material well-being (MW), self-determination (SD) and rights (RI). These eight dimensions cover three essential areas, the personal, the social, and the judicial area of life.

B. Intellectual disability in older people

The requirement for age-specific support is made more urgent by the longer life expectancy of those with ID. Generally, elder people with ID live in residential services. Where caregivers give the best possible support to them to make their life easier. However, sometimes caregivers are challenged by their evolving requirements as they strive to provide the highest level of care [12]. On the other hand, ID people face many challenges in their personal and professional lives. Throughout their progressively longer lives, people with intellectual disabilities experience many unfavourable life events. According to reports, people with intellectual disabilities endure more traumatic, potentially upsetting life situations than average, not just when they're young but also throughout their much longer lifespans. The prevalence of violent crime, abuse, neglect, and other forms of discrimination ranges from 30% to approximately 90% [13]. Such events may result in toxic stress, which then causes trauma. Trauma happens when a person's resources cannot handle stress [14].

C. The GENCAT tool

The GENCAT tool is proposed by Verdugo et al. [8]. The GENCAT scale is used to gauge beneficiaries' overall quality of life. This scale is created as a measurement tool to support social services' ongoing improvement. The multidimensional model on which GENCAT has based displays individual wants regarding eight key dimensions (EW, IR, PW, PD, MW, SD, SI, RI) [9]. The scale measures the quality of life using 69 questionnaires that all responded on a 4-point Likert scale, all declaratively presented in the third person. There are various scales in the GENCAT Scale depending on the type of population being studied: a general population scale, an elderly population scale (those who are 50 years of age or over), a scale for intellectually disabled persons, and a scale for other groups (e.g., physical disability, drug addicts, and mental health problems). We applied the scale for intellectually disabled issues [10].

D. ML-based models

ML models are prone to learn from structured data. ML models are categorised into three categories: Supervised, unsupervised, and reinforcement learning. The nature of our dataset is structured and labelled dataset. Data has 69 input features and corresponding index values as output labels. Therefore, supervised learning algorithms are the best fit for this dataset. In supervised learning, there are two categories: regression and classification. These two categories rely on the nature of output features. If the output feature is categorical, then classification algorithms are helpful, whereas, in the case of regression, the nature of the output variable will be some numeric value that relies on the input features value. We use regression algorithms based on the nature of our data, where the output feature is an index value.

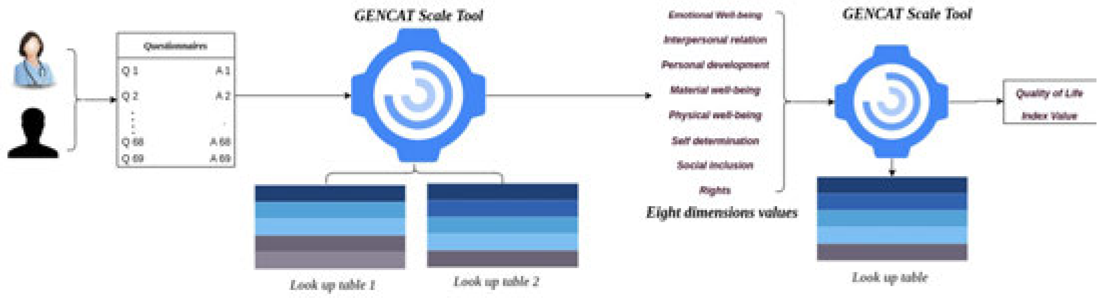

Fig. 1. Figure shows the complete architecture to predict the quality of life index value using the Gencat tool. It starts with an interview conducted between the professional and the beneficiary. Professionals collect the response to the 69 questionnaires, and using the Gencat tool, he/she convert the response into the eight dimensions value. After that, using the eight-dimension value, professionals calculate the quality of life index value using the Gencat tool.

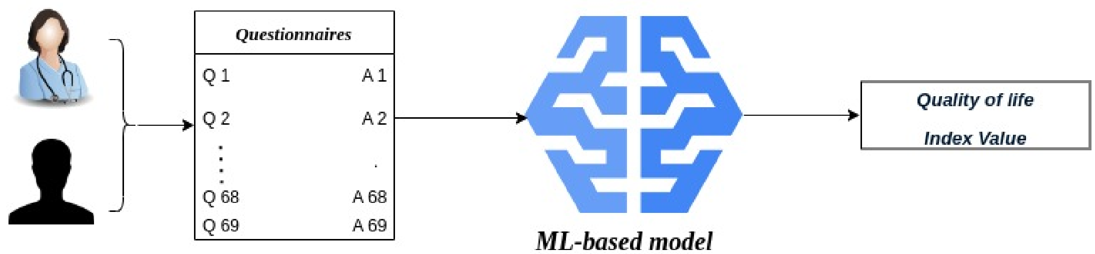

Fig. 2. Figure shows the complete architecture of the proposed method. After the interview, a trained ML-based model takes the response of 69 items as input and predicts the quality of life index value.

III. Methodology

This work proposes to use an ML-based model to predict the QoL index value directly from the response of the 69 items. Earlier, people used the GENCAT tool to calculate the index value, which is a traditional and statistical-based method where many rules and relatable tables are involved in calculating the first eight dimensions value using 69 questionnaires response and after that, using the eight dimensions values, the professional again calculate the QoL index value using the GENCAT scale as shown in Fig. 1. This process of calculating the QoL index value from the 69 items response is very skills-required and needs a professional to calculate these values. We propose a learning-based approach to directly predict the QoL index value by taking 69 questionnaires response as shown in Fig. 2. We trained various regression-based algorithms using the NewtonOne dataset. We assessed the models' performance using error metrics, including mean squared error, mean absolute error, root mean squared error and R2 score value.

A. NewtonOne dataset

The NewtonOne dataset is a private dataset collected by professionals during Never Alone project. This NeverAlone project is ongoing. Therefore, the number of data is increasing time after time. Currently, this data contains QoL data of 113 persons. The category of people who participated in this data collection has a mild ID and an age group above 55. The demography of the data has 65% female participants and 35% male participants. One data contains 69 input features and corresponding QoL index values. The response of each 69 questionnaires is collected using four points Likert scale. Professionals during the interview asked questions to the participants and, based on the answers, gave the value as a response.

B. Data augmentation

Data is the backbone of machine learning models. For training an ML-based model, the amount of data plays an essential role in the model's performance. Therefore, a considerable amount of data is required to train an ML-based model for optimal performance. The data augmentation technique is used to increase the number of data. Therefore, this technique is commonly used in case of fewer data. This work uses the NewtonOne data, which contains 113 original data points. We trained various ML-based; therefore, increasing the data as much as possible is demanded.

Augmentation is used to train the model to increase its performance. Whereas the model evaluation requires the original data to justify the model's performance. Therefore, we split our dataset into train and test sets before augmentation. The ratio of the train and test is set as 70% and 30%. After that, we augmented the training dataset. Using the augmented train dataset, we train the ML-based and DNN models and test the model's performance using an original test set data. In this experiment, we imported the ML algorithms using the Scikit-learn library [15]. Implementation of the experiment done on Google Colaboratory [16].

SMOGN [17] (Synthetic minority over-sampling technique for regression with random Gaussian noise) is proposed to deal with the imbalanced distribution regression task. SMOTE-R [18] and adding Gaussian Noise are two oversampling techniques used with random under-sampling in SMOGN.

When a high risk is associated with using SMOTE-R to generate a new example, a more cautious alternative is to generate the new example by adding Gaussian noise. The distance between the base scenario and the neighbour case, to be more precise, is what determines which approach to utilize. SMOTE-R is employed if the neighbour case is close enough; otherwise, Gaussian Noise generates the new case. Additionally, SMOGN resamples the dataset using both over and under-sampling [19].

C. Quality of life index value

The quality of life index value is essential in providing support to intellectually disabled people [20]. This value decides whether a person needs support or not. Therefore the role of the index value is very important in providing support. The index value follows the Gaussian nature, and the 80% of the population lies between the Gaussian bell-shaped curve, whose values vary from a minimum of 68 to 130. Most of the population will come under these values. The mean of the curve lies at 100. therefore if the index value of a beneficiary is 100 or above 100, then he/she does not need any support. If the index value is below 100, she/he needs support. Professionals use the GENCAT tool to calculate the index value in steps, where they calculate the first 8 dimension values from the 69 questions and then calculate the index value using the 8 dimensions value. In this work, we propose to use an ML-based model to predict the index value directly using 69 items responses.

D. Developing ML-based QoL prediction models

Even though classic machine learning techniques have made tremendous progress in knowledge discovery, they sometimes struggle to perform well when faced with complex data, such as unbalanced, highly dimensional, noisy data, etc. The cause of this is that various qualities and the underlying structure of the data are complex for such methods to represent [21]. A set of features are first extracted through ensemble learning using various transformations. Various learning algorithms are used to get mediocre predicted outcomes based on these learned properties. Compared to a single model, this strategy enables more excellent predictive performance.

Bayesian regression enables a natural process to persist in sparse or uneven data. Instead of using point estimates, it formulates linear regression utilizing probability distribution. It is anticipated that the result will be chosen from a probability distribution instead of being evaluated as a single value [22]. Regularized linear regression is the basis of the Bayesian algorithm. The approach uses a hyperparameter to adjust regularisation strength to fully integrate over the hyperparameter inside the posterior distribution and apply a roughly noninformative hyperprior. Because Bayesian linear regression includes Bayesian ridge regression and belongs to the family of ridge regressions, and exhibits all the characteristics of ridge regression and Bayesian linear regression [23].

Regularization and feature selection are two critical tasks that can be accomplished with the help of the Least Absolute Shrinkage and Selection Operator (Lasso) [24]. The LASSO method constrains some of the model's absolute parameter values. The total must fall below a certain threshold (upper bound). To achieve this, the method employs a shrinkage (or regularization) mechanism that penalizes the regression variables' coefficients, therefore decreasing some of them to zero. During the feature selection stage, the variables with non-zero coefficients that the shrinking technique left are picked to be a component of the model. The method's objective is to reduce prediction error [24].

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(Y_i - \hat{Y}_i)^2 + \lambda\sum_{j=1}^{n}|\beta_j|, \tag{1}$$

where $\frac{1}{m}\sum_{i=1}^{m}(Y_i - \hat{Y}_i)^2$ is the mean squared loss function and $\lambda\sum_{j=1}^{n}|\beta_j|$ is the penalty $L1$ regularization.

Ridge regression is a refinement of conventional least squares that extends the optimization issue by including a regularization term [25]. The regularization hyperparameter in ridge regression needs to be determined from the data. The model will tend to overfit the data noise if the estimated value is too low, and it will not forecast as well as it might if it is too high. Usually, the regularization parameter is chosen using the grid search with a cross-validation approach.

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(Y_i - \hat{Y}_i)^2 + \lambda\sum_{j=1}^{p}{\beta_j}^2, \tag{2}$$

where $\frac{1}{m}\sum_{i=1}^{m}(Y_i - \hat{Y}_i)^2$ is the mean squared loss function and $\lambda\sum_{j=1}^{p}{\beta_j}^2$ is the penalty $L2$ regularization.

IV. Results

A. Evaluation metrics

The model's performance is measured using error metrics in the regression task. These error metrics are mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE) and R^2 score.

MAE measures the mean of the absolute difference between the ground truth value and the model's predicted value.

$$MAE = \frac{\sum_{i=1}^{n}|Y_i - \hat{Y}_i|}{n}, \tag{3}$$

here, $\hat{Y}_i$ is the predicted value and Y_i is the ground truth value. i varies from 1 to n.

The standard deviation of the residuals is RMSE. Residuals and RMSE measure the spread of these residuals, which is the distance between the data points and the regression line. To put it another way, it indicates how closely the data is centred on the line of best fit. RMSE is commonly used in regression analysis, forecasting, and climatology to verify the accuracy of experimental results.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}}, \tag{4}$$

where, Y_i is the actual value and $\hat{Y}_i$ is the predicted value.

R-squared measures a linear regression model's ability to fit data. This statistic demonstrates the extent to which the

TABLE I
TEST CASE RESULTS OF ORIGINAL DATA FOR DIFFERENT ML ALGORITHMS

No	ML Models	MAE	RMSE	R^2 score	Variable parameters
1	Extra trees regressor (ETR)	5.4073	48.7362	0.6785	num of estimators= 20
2	Random forest regressor (RFR)	5.1991	6.9963	0.6771	$maxdepth = 15$
3	Gradient boosting regressor (GBR)	5.7362	8.2507	0.5509	
4	K neighbors regressor (KNN)	7.4926	9.1553	0.4471	
5	Bayesian ridge (BR)	1.7648	2.1664	0.9690	
6	Linear regression (LR)	2.3722	3.1927	0.9327	
7	Lasso regression (Lasso)	1.9504	2.4140	0.9615	$alpha = 0.01$
8	Ridge regression (Ridge)	1.6667	1.9643	0.9745	
9	Huber regressor (Huber)	4.3128	5.4135	0.8066	
10	Elastic net (EN)	3.1787	3.9597	0.8965	
11	Orthogonal matching pursuit (OMP)	7.0742	8.4859	0.5250	
12	AdaBoost regressor (Ada)	5.4395	6.7094	0.7030	num estimators=300
13	Decision tree regressor (DT)	10.1764	14.2663	-0.3425	
14	Least angle regression (LAR)	5.2143	5.9376	0.7674	num nonzero coefs = 30
15	Passive aggressive regressor (PAR)	5.7502	6.7163	0.7024	

TABLE II
TEST CASE RESULTS AUGMENTED DATA FOR DIFFERENT ML ALGORITHMS

No	ML Models	MAE	RMSE	R^2 score	Variable parameters
1	Extra trees regressor (ETR)	5.8897	7.2597	0.6865	num estimators=20
2	Random forest regressor (RFR)	5.9216	7.1776	0.6936	max depth=15
3	Gradient boosting regressor (GBR)	5.1671	6.1371	0.7760	
4	K neighbors regressor (KNN)	7.2669	8.7813	0.5414	
5	Bayesian ridge (BR)	2.0681	2.7359	0.9554	
6	Linear regression (LR)	3.8803	5.3497	0.8298	
7	Lasso regression (Lasso)	2.0397	2.5516	0.9612	$alpha = 0.01$
8	Ridge regression (Ridge)	1.5103	1.9956	0.9763	
9	Huber regressor (Huber)	4.1738	5.5782	0.8149	
10	Elastic net (EN)	2.6404	3.7010	0.9185	
11	Orthogonal matching pursuit (OMP)	11.0497	12.8025	0.0253	
12	AdaBoost regressor (Ada)	5.7854	7.5685	0.6593	num estimators=300
13	Decision tree regressor (DT)	10.7058	13.1507	-0.0284	
14	Least angle regression (LAR)	4.6589	5.8079	0.7994	num nonzero coefs=30
15	Passive aggressive regressor (PAR)	3.4341	4.7778	0.8642	

independent variables may collectively explain the variation of the dependent variable. R-squared concisely measures the strength of the relationship between the model and the dependent variable using a useful 0–100% scale.

$$R^2 Score = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y}_i)^2}, \tag{5}$$

where, $\hat{Y}_i$ is the predicted value, Y_i is the actual value, $\bar{Y}$ is the mean value.

B. Ablation study

We experimented with two ways based on the availability of the data. We trained the model using the augmented and original data and reported the results for both in the tables I and II

Table I shows the various ML-based model performance for the test case result of the original dataset. The original dataset contains 113 beneficiary's responses to 69 items and corresponding index values. We split the original data into 70% (79) for training and 30% (34) for testing. The results are shown in the form of four evaluation metrics. The variable parameters show a change in variable values to the default value. In other models' cases, we directly imported the model and did not change their variable because we achieved the best results from the default value, whereas in other cases, we achieved better results when we used the defined variable by us.

Table II depicts the various ML-based model performance for the augmented data. Initially, the data contained a total of 113 beneficiary information. Before augmentation of the original data, we split the data into train and test in the ratio of 70% and 30%. The train data contains 79 data, and we use the SMOGN algorithm to augment the train data. Whereas we do not augment the test data. After the augmentation, the total data point in the data increases to 150. So, in this case, we trained various ML-based models using 150 data and tested the model's performance using 34 data.

Tables I, II show the results of the two case data of the various ML-based model in the form of MAE, RMSE, and R^2 score value. We imported various ensemble based ML-models to learn the correlation between the 69 questionnaire responses and the corresponding index value. For both cases, the performance of the three models is higher than the others. Bayesian ridge, Lasso regression and Ridge regression have the minimum error and the highest R^2 score value. The evaluation metrics score for Bayesian ridge for the test case is $1.7648, 2.1664$, and 0.9690 as MAE, RMSE, and R^2 scores, respectively, in the case of original data. The results of Lasso regression for the test case are $1.9504, 2.4140$, and 0.9615

as MAE, RMSE, and R^2 scores, respectively, in the case of original data. And the result of the Ridge regression for the test case scenario is $1.6667, 1.9643$, and 0.9745 as MAE, RMSE, and R^2 scores, respectively, in the case of original data. The performance of these three models on original data is as good as or even better than augmented case results. The performance of the other models, like random forest regressor, gradient boosting, passive-aggressive regressor and others, is better in the augmented data case. The Ridge regression model performance is superior to all other 14 models in both cases. The ridge regression results in augmented data cases are $1.5103, 1.9956$, and 0.9763 as MAE, RMSE, and R^2 scores, respectively.

The performance of the various model depends on the number of the dataset. In our case, we have limitations in the original dataset. And one can not augment the data more because the augmented data may lose the originality of the distribution. Therefore, an increasing number of original data can improve the model's performance.

V. Conclusion

This paper proposed a novel approach for predicting the QoL index, and it solves the limitations of the existing manual approach. We use the NewtonOne dataset to train the various state-of-the-art machine learning models. So far, professionals have used the GENCAT scale to calculate the QoL index by taking first the response of the 69 questionnaires and, after that calculating the eight dimensions value and from eight dimensions value to the QoL index value. Which is a lengthy and tedious task and requires a skilful professional to calculate the index value. We propose using a trained ML-based model which accurately predicts the QoL index value by inputting the response of the 69 questionnaires. In the traditional approach calculating the index value is a two-step process in which first calculates the eight dimension value from the 69 questionnaires response and after that from eight dimension value to the QoL index value. In this approach, we directly calculate the index value. Out of the various regression model, the performance of the Ridge regression model is superior in terms of the MAE, RMSE, and R^2 score value. Therefore we finalize using this algorithm to predict the QoL index value.

We will incorporate more data in the NewtonOne dataset in future work to improve our model performance and add sensor-based information affecting the QoL of the individual together with questionnaires response.

References

[1] G. L. Krahn, "Who world report on disability: a review," *Disability and health journal*, vol. 4, no. 3, pp. 141–142, 2011.

[2] D. R. Patel, M. D. Cabral, A. Ho, and J. Merrick, "A clinical primer on intellectual disability," *Translational pediatrics*, vol. 9, no. Suppl 1, p. S23, 2020.

[3] K. McKenzie, M. Milton, G. Smith, and H. Ouellette-Kuntz, "Systematic review of the prevalence and incidence of intellectual disabilities: current trends and issues," *Current Developmental Disorders Reports*, vol. 3, no. 2, pp. 104–115, 2016.

[4] D. Wieczorek, "Autosomal dominant intellectual disability," *medizinische genetik*, vol. 30, no. 3, pp. 318–322, 2018.

[5] L. F. Oliveira, T. F. Chaves, N. Baretto, G. R. de Luca, I. T. Barbato, J. H. Barbato Filho, M. Ocampos, and A. F. Maris, "Etiology of intellectual disability in individuals from special education schools in the south of brazil," *BMC pediatrics*, vol. 20, no. 1, pp. 1–12, 2020.

[6] I. Grabowska, R. Antczak, J. Zwierzchowski, and T. Panek, "Individual quality of life and the environment–towards a concept of livable areas for persons with disabilities in poland," *BMC public health*, vol. 21, no. 1, pp. 1–15, 2021.

[7] G. K. Yadav, B. M. Vidales, H. A. Rashwan, J. Oliver, D. Puig, G. Nandi, and M. Abdel-Nasser, "Effective ml-based quality of life prediction approach for dependent people in guardianship entities," *Alexandria Engineering Journal*, 2022.

[8] M. Á. Verdugo, B. Arias, L. E. Gómez, and R. L. Schalock, "Development of an objective instrument to assess quality of life in social services: Reliability and validity in spain," *International Journal of clinical and health psychology*, vol. 10, no. 1, pp. 105–123, 2010.

[9] G. K. Yadav, B. M. Vidales, S. Dueñas, M. Abdel-Nasser, H. A. Rashwan, D. Puig, and G. Nandi, "Predicting personalized quality of life of an intellectually disabled person utilizing machine learning," in *Artificial Intelligence Research and Development*. IOS Press, 2022, pp. 139–142.

[10] A. Maestro-Gonzalez, M. C. Bilbao-Leon, D. Zuazua-Rico, J. M. Fernandez-Carreira, R. F. Baldonedo-Cernuda, and M. P. Mosteiro-Diaz, "Quality of life as assessed by adults with cerebral palsy," *PLoS One*, vol. 13, no. 2, p. e0191960, 2018.

[11] L. E. Gomez, B. Arias, M. A. Verdugo, and P. Navas, "An outcomes-based assessment of quality of life in social services," *Social Indicators Research*, vol. 106, no. 1, pp. 81–93, 2012.

[12] H. R. Schepens, J. Van Puyenbroeck, and B. Maes, "How to improve the quality of life of elderly people with intellectual disability: A systematic literature review of support strategies," *Journal of applied research in intellectual disabilities*, vol. 32, no. 3, pp. 483–521, 2019.

[13] N. J. Baladerian, T. F. Coleman, and J. Stream, "Abuse of people with disabilities," *Spectrum Institute Disability and Abuse Project*, 2013.

[14] A. S. Masten and A. J. Barnes, "Resilience in children: Developmental perspectives," *Children*, vol. 5, no. 7, p. 98, 2018.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[16] T. Carneiro, R. V. M. Da Nóbrega, T. Nepomuceno, G.-B. Bian, V. H. C. De Albuquerque, and P. P. Reboucas Filho, "Performance analysis of google colaboratory as a tool for accelerating deep learning applications," *IEEE Access*, vol. 6, pp. 61 677–61 685, 2018.

[17] P. Branco, L. Torgo, and R. P. Ribeiro, "Smogn: a pre-processing approach for imbalanced regression," in *First international workshop on learning with imbalanced domains: Theory and applications*. PMLR, 2017, pp. 36–50.

[18] L. Torgo, R. P. Ribeiro, B. Pfahringer, and P. Branco, "Smote for regression," in *Portuguese conference on artificial intelligence*. Springer, 2013, pp. 378–389.

[19] X. Y. Song, N. Dao, and P. Branco, "Distsmogn: Distributed smogn for imbalanced regression problems," in *Fourth International Workshop on Learning with Imbalanced Domains: Theory and Applications*. PMLR, 2022, pp. 38–52.

[20] M. Verdugo, P. Navas, L. Gómez, and R. L. Schalock, "The concept of quality of life and its role in enhancing human rights in the field of intellectual disability," *Journal of Intellectual Disability Research*, vol. 56, no. 11, pp. 1036–1045, 2012.

[21] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers of Computer Science*, vol. 14, no. 2, pp. 241–258, 2020.

[22] S. M. Mostafa, A. S. Eladimy, S. Hamad, and H. Amano, "Cbrl and cbrc: Novel algorithms for improving missing value imputation accuracy based on bayesian ridge regression," *Symmetry*, vol. 12, no. 10, p. 1594, 2020.

[23] Y. Yang and Y. Yang, "Hybrid prediction method for wind speed combining ensemble empirical mode decomposition and bayesian ridge regression," *IEEE Access*, vol. 8, pp. 71 206–71 218, 2020.

[24] S. Shafiee, L. M. Lied, I. Burud, J. A. Dieseth, M. Alsheikh, and M. Lillemo, "Sequential forward selection and support vector regression in comparison to lasso regression for spring wheat yield prediction based on uav imagery," *Computers and Electronics in Agriculture*, vol. 183, p. 106036, 2021.

[25] T. D. la Tour, M. Eickenberg, A. O. Nunez-Elizalde, and J. L. Gallant, "Feature-space selection with banded ridge regression," *NeuroImage*, vol. 264, p. 119728, 2022.