CHAPTER 3
IDENTIFICATION OF SOCIO-ECONOMIC INDICATORS USING MIXED-METHOD APPROACH OF PCA AND FUZZY-TOPSIS

3.1 Introduction

Witnessing limited redistributive channels, governments of developing countries and aid donor agencies look forward in achieving distributional objectives by facilitating income opportunities, basic health, and educational facilities among the rural population through road interventions. Road interventions facilitate efficiency benefits by encouraging rural households in taking up new opportunities, which help them in relieving their well-being constraints.

Globally road systems are recognized as the significant contributors for the economic and social development of a nation. In the case of rural areas, roads improve mobility as well as access to basic services and market centres. In recent times, most of the developing countries have given emphasize on improving road infrastructure for rural areas, and India is no exception. Government of India initiated a major rural road development plan in the country, known as Pradhan Mantri Gram Sadak Yojana (PMGSY), in the year 2000. The objective of this program is to connect rural habitations having a population of 250 persons and above (for desert and hilly regions) by all-weather roads to the nearest village or market centres. The scheme is targeted for poverty alleviation and development of rural habitations thereby enhancing their socio-economic status.

Improvised road infrastructure in context with rural areas has a significant impact on the target population. They provide new avenues and employment opportunities for rural inhabitants (Riverson et al., 1991) and bring out economic growth with poverty alleviation (Banister and Berechman, 2003). Rural roads generate market activities due to reduced transport costs, as well as foster linkages to economic centres which help rural habitants to enhance their agricultural production. Rural roads also stimulate non-farm activities along with alteration in land use, crop diversification (Van de Walle, 2009). Better rural roads enhance social outcomes by facilitating access to social services such as education and health facilities. This is actualized in terms of increase in the number of school going children due to the reduction in travel time to reach the facility (Khandker et al., 2009). The same holds true in case of access to a health facility;

individuals can get treatments at the first call due to good road connectivity (Rocha and Soares, 2010).

Despite of consensus on how rural roads are important in the development of living standards of the rural population, surprisingly there is little evidence in the literature which captures the size and nature of benefits in a comprehensive way. Indeed, there are few rigorous studies which assess the benefits of rural roads in credible manner, but they still lag to capture distributional impacts induced by them. Traditionally, planning of roads and their investment decisions have been prioritized based on cost-benefit assessments. The studies attempted to assess rural road investment by considering the savings incurred in terms of vehicle operating cost and reduction in travel time. However, in case of developing countries where the traffic in rural areas is too low, application of conventional methods such as cost and benefit analysis cannot be relied completely (Van de Walle, 2002).

Moreover, rural road infrastructure is need based, i.e., they are constructed not just for the sole purpose of travel but also to improve the socio-economic condition of the target population. At the same time, the target population served by these roads is diverse in terms of socio-economic backgrounds with different necessities, which makes the task of assessment of rural road investment complex. Also, some of the impacts may be direct or indirect (positive or negative) which are difficult to be captured by using conventional cost-benefit analysis (Grootaert and Calvo, 2002). The traditional cost-benefit analysis is therefore required to be modified by taking into consideration of different factors associated with socio-economic impact assessment methodology (SEIA), so that expected benefits and costs of the different groups are measured in comprehensive way. Assessment of impacts adds up as an input to the decision makers by providing better information on both positive and negative impacts of delivered infrastructure. Impact evaluation is an important tool in policy-oriented executions (Ehrlich and Ross, 2015).

In recent times, studies have been performed to understand the impacts of rural roads construction using different impact evaluation techniques. However, it has been observed that these studies faced difficulties in assessing the magnitude of the impacts due to the underlying problem of endogeneity as well as identification of proper indicators, which assist the impact evaluation process in a comprehensive manner (Binswanger et al., 1993; Jalan and Ravallion, 1998; Rowan, 2012). Most common shortcomings of previous studies are, selection of appropriate indicators and the target population, which are influenced by the placement of the rural roads as well as its outcomes. Better evaluation process of the impacts requires proper identification of the indicators (i.e., data), which are of potential importance and are affected directly or indirectly by the improvements in the roads.

In the road impact evaluation process, first step is to know about the kind of the impacts rural roads can instigate on the target population. Further, it requires proper understanding and knowledge about the potential indicators which will be impacted the most. For example, in what way, initial conditions of the rural households interact with roads to influence their outcomes. Next is to identify the target population which is then followed by ascertaining of proper evaluation technique. Finally, the magnitude of these impacts is determined in terms of the criticality of the influential indicators so that appropriate methodologies can be developed to address design and selection problems by taking into consideration of rigorous impact evaluations for road project appraisals. Thus, keeping this in view, this study develops a novel methodology to explore and ascertain important indicators by considering ex-post evaluation condition for newly constructed all-weather rural roads.

The novel methodology of the present study is based on the concept of mixed method design (i.e., concurrent triangulation design), where the findings of both quantitative and qualitative techniques are compared to cross-validate the outcome. In the present study, principal component analysis (PCA) which considers quantitative assessment is compared with fuzzy multi-criteria decision making (MCDM) which contemplates qualitative assessment. A case study of rural roads constructed under the PMGSY scheme in Jhunjhunu district of Rajasthan state of India is considered to illustrate the effectiveness of the proposed methodology. In order to set the flow of the study, the chapter is divided into five sections.

Section 3.1 introduces the study and its need to be conducted. It also discusses some of the generic issues that curtail the selection process of indicators in context with impacts instigated due to improvements in rural roads. Section 3.2 then addresses the methodology by addressing the case study along with criteria/indicator selection and data collection. Next, Section 3.3 discusses the proposed methodological approach as well as the steps followed in the evaluation. Section 3.4

briefly discusses the results and findings of the proposed methodology, while Section 3.5 summarizes the study by considering the assessment and findings of the proposed study.

3.1.1 Significance of rural road impact assessment indicators

Appropriate impact indicators are used to measure the goal and the targets of the delivered infrastructure and form the basis of sustainability planning and comprehensive management of road infrastructure. Impact indicators play a vital role in establishing baseline of impacts and help in identifying their trends. They can significantly influence the assessment process if not selected appropriately. Thus, selecting them pose a vital challenge as they provide useful information on the goals achieved by the delivered road infrastructure. As, rural roads instigate various impacts on social, economic, and environmental aspects of the target population. Many a times, employing a single indicator to assess the impact is not adequate; rather it can be addressed in a better way by a set of indicators. Thus, allowing for comprehensive assessment, i.e., economic impacts can collectively be well defined by the indicators such as an increase in individual/household income, availability of jobs, income diversification, and even few more indicators (Litman, 2007).

Indicators selected, are considered to have many uses: they can help in identifying possible trends, predict impacts, assess intervention based on performance goals, and evaluate the effects of the intervention on an identified section of the population (target population). Therefore, it is important to select indicators carefully that reflect the overall aim of the scheme or intervention. Also, indicators selected are often required to be realistic from the viewpoint of availability of data, their ability to perceive and convenience in decision making. Hence, it is necessary to understand the perspectives and limitations of each indicator. However, there exist a tension between suitability and completeness while identifying indicators. If small indicators set is selected, it may overlook important impacts. It possibly may distort overall outcome, though convenient to use considering the availability of data, whereas, a larger set may not be cost-effective and will be difficult to quantify (Morimoto, 2013).

Currently, no standardized techniques are available, to identify reliable set of indicators for a comprehensive assessment of road infrastructure impacts. They are developed based either on the needs or abilities. Therefore, this creates a need to develop a methodology that would be useful in establishing recommended transportation indicator sets from the viewpoint of sustainable planning and evaluation best practices.

3.2 Methodology

3.2.1 Case study

The present study follows the ex-post approach in selecting road stretches and identifies them based on their geographical location and the size of the target population. The study also considers the year of their execution in selecting road stretches (i.e., a newly laid road does not instigate impacts immediately, except a few benefits. On the other hand, when road ages and becomes a part of the structure in the village, it is difficult for the inhabitants to appreciate the impacts). The aim is to have a dataset with reduced errors (biases associated with the perception of rural inhabitant's). Thus, considering these aspects, present analysis selects road stretches in six different blocks, viz., Buhana, Jhunjhunu, Khetri, Surajgarh, Nawalgarh, and Udaipurwati of Jhunjhunu district in Rajasthan state, India. The road stretches have been constructed in the year 2013-2014 under PMGSY scheme. A total of 27 new connectivities are considered for the assessment with the population of 9640 persons who have been directly served by them. Fig. 3.1 shows the number of blocks in Jhunjhunu district along with through routes employed for the study.

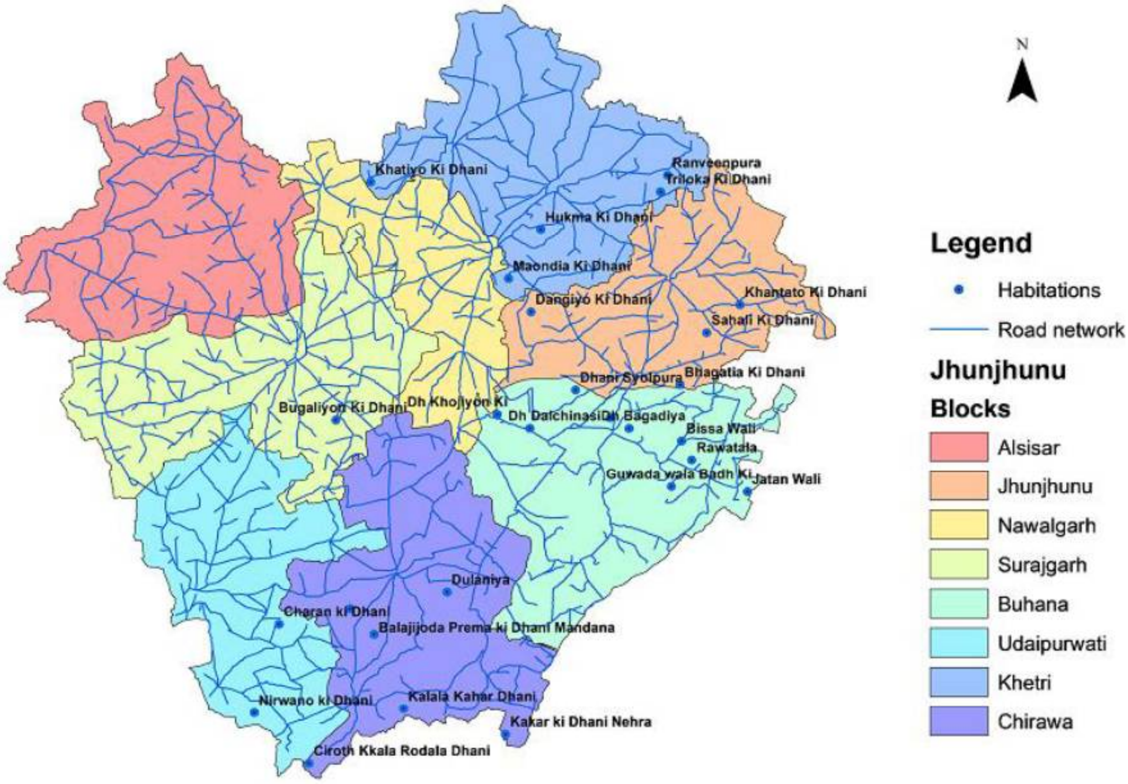

Fig. 3.1 Block boundaries showing through routes in Jhunjhunu district

3.2.2 Selection of criteria/indicator

Assessing rural road projects and their impacts is an important aspect from the viewpoint of the welfare of the community. The impacts of construction of rural roads are many, and some of them are not attained immediately (i.e., they are immediate, mid-term and long-term). Moreover, many of the impacts cannot be quantified. Thus, the selection of appropriate criterion for assessing the impacts must be done carefully. The selected criteria should account the change which is both qualitative and quantitative in nature and are required to be competent enough to account the impacts comprehensively. In consideration of these aspects, the study follows a systematic method in selecting important criteria as illustrated in Fig. 3.2. Initially, the study focuses on the available scientific literature which is followed by opinions from the expert group. The expert group consisted of a team of five members belonging to educational and research institutes, authorities working in the field of rural development schemes in the government organizations. Moreover, a preliminary survey for a few selected habitations has been done to get direct feedback from the rural inhabitants and accordingly a concise set of 33 sub-criteria defining five main criteria/indicators is finalized. Table 3.1 below depicts the SEIA criteria considered for the study.

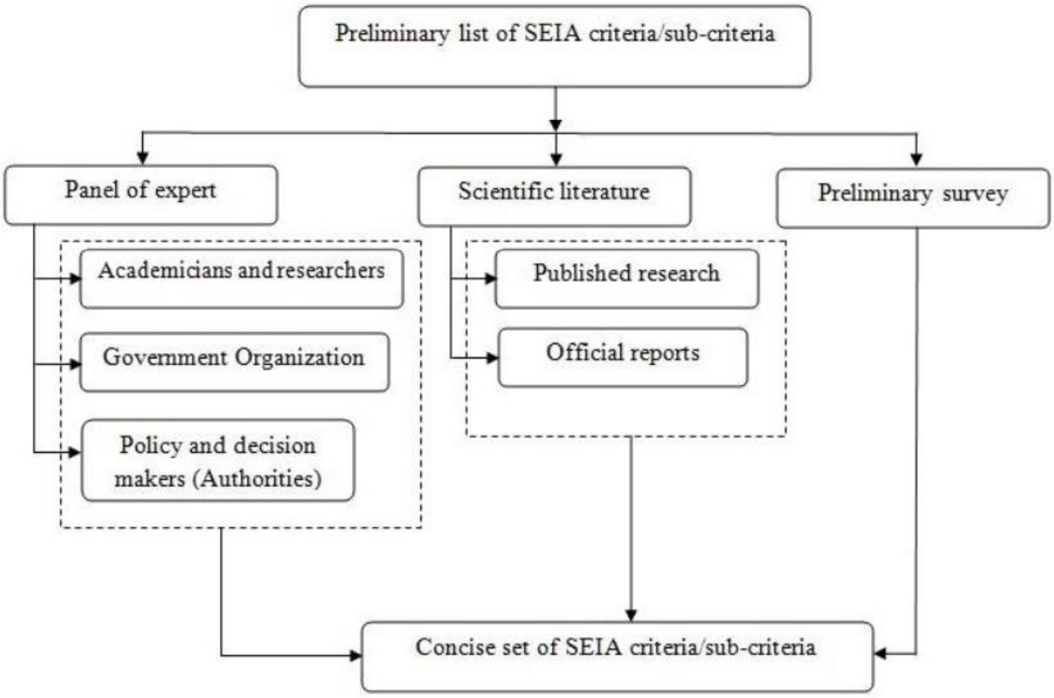

Fig. 3.2 Process followed for SEIA criteria selection

Table 3.1 SEIA criteria with sub-criteria employed in the study

Criteria/Indicators				
Transport Facility	Income Status	Health Facility	Education facility	Quality of Neighborhood (social environment)
Travel time using public transportation (T_{TTPUB})	Individual Income (I_{II})	Use of health facility (H_{FU})	Literacy rate of male (E_{MLR})	Ownership of personal phone (Q_{PPO}).
Travel time using private	Household income (I_{HH})	Availability of health clinic (H_{CA})	Literacy rate of female (E_{FLR})	Ownership of Television (Q_{TVO})

transportation (T_{TTPVT})				
Public transportation units (T_{PUBTU})	Income of self-employed from agriculture (I_{SA})	Availability of primary health center (H_{PHCA})	Percent of male children attending schools (E_{MAS})	Livability (Q_L)
Private transportation units (T_{PVTTU})	Income of wage labor from agriculture (I_{WA})	Access to the mode of transport for health facility (H_{HAM})	Percent of female children attending schools (E_{FAS})	Involvement in Social-gathering within the village (Q_{SGIV})
Frequency of Public transportation (T_{PUBF})	Income from Livestock (I_L)	Travel time to reach health facility (H_{TT})	Access to the mode of transport for Education facility (E_{AM})	Involvement in Social-gathering outside the village (Q_{SGOV})
Public transportation cost (T_{PUBTC})	Income of unskilled labor from agriculture (I_{UA})	Health Status (anthropometric measures up to adolescent age) (H_{HSANT})	Travel time to reach education facility (E_{TT})	
Private transportation cost (T_{PVTTC})	Income of unskilled labor from non-agriculture (I_{UNA})		Availability of Preschools (E_{PRESA})	
			Availability of Primary schools (E_{PRISA})	

3.2.3 Data collection

Collection of data and its assessment are essential in the decision-making process as well in predicament elucidation (Hair et al., 1995). One of the key mechanisms to be employed for collecting necessary data for impact assessment study is through focus-group survey (perceptions of the target population). The data for the present study is collected from 27 habitations connected by PMGSY roads in Jhunjhunu district of Rajasthan State, India. The sample size required for data collection is 370 persons, considering total population (9640 persons) served by these habitations and with 95% confidence interval and 5% of marginal error. Thus, 27 Focus-groups @14 participants per group were identified from the respective habitations. The participants consisted of government and private employees, self-employed (farmers and traders) and students (age 16 to 45 years). The focus group consisted of 66% of male and 34% of female participants. A preliminary survey has been conducted before the final survey to avoid potential risks (indulgence of error) associated with the overall survey process.

The focus group surveys have been conducted in the month of April–May 2016. All discussions are based on a questionnaire, designed after a comprehensive study and is broadly divided into five sections, viz., impacts on transport facility, income status, education facility, health facility and quality of neighbourhood (social environment). The perceptions of focus groups are collected to capture the necessary information required on how the criteria have impacted the inhabitants, and to consider their level of satisfaction. The level of satisfaction captured for each of the indicators is gauged on a linguistic scale ranging from highly satisfied to extremely dissatisfied. The scores are assigned from 1 to 5 (5 being highly satisfied and 1 being not acceptable). Before commencement of formal data collection, the enumerators had a general discussion with the participants about the habitation and their lifestyle. This facilitated the formal data collection process with ease and comfort between enumerators and participants. The authenticity of data has been ensured with participants through feedback at the end of group discussions.

3.3 Methods

3.3.1 Mixed method approach

Mixed method approach is a technique which combines quantitative and qualitative approaches or concepts. It combines methodological approaches considering their fundamental aspects. As, value addition they have the ability to address the problem by considering various viewpoints so that proper comprehension about the problem assessment is achieved. Application of mixed method research has increased considerably (Creswell, 2006; Schoonenboom and Johnson, 2017) and has motivated researchers to move beyond the argument between quantitative and qualitative techniques (Morgan, 2007; Mele and Belardinelli, 2018). Despite its usefulness, it poses challenge, such as how to design overall methodology, whether both quantitative and qualitative methods are to be given equal priority or to be used concurrently or sequentially and how to integrate them. Thus, keeping in view of the above facts, the present study primarily focuses on the overall design and interactions of both the quantitative and qualitative methods.

The present study is motivated by the designs proposed by Leech and Onwuegbuzie (2009). Initially, sequential mixed method design has been considered, but there has been difficulty in having proper elucidation in reference with the objective of the study, therefore concurrent triangulation mixed method approach has been considered. The aim has been to rely on the outcomes of the quantitative method and use qualitative assessment techniques to supplement and complement it by validating the assessment process. Thereby, allowing us to improve our assessment objective. In the present study, a novel concurrent triangulation design of mixed methods research (MMR) has been proposed, for exploring and ascertaining important indicators by considering ex-post evaluation condition for newly constructed all-weather rural roads. It integrates both qualitative and quantitative assessment approaches which provide complete understanding about the effects which can yield generalized outcome when applied singularly.

In the present study, principal component analysis (PCA) which considers quantitative assessment is compared with fuzzy multi-criteria decision making (MCDM) which contemplates qualitative assessment. The study procedure is chronologically outlined in Fig. 3.3.

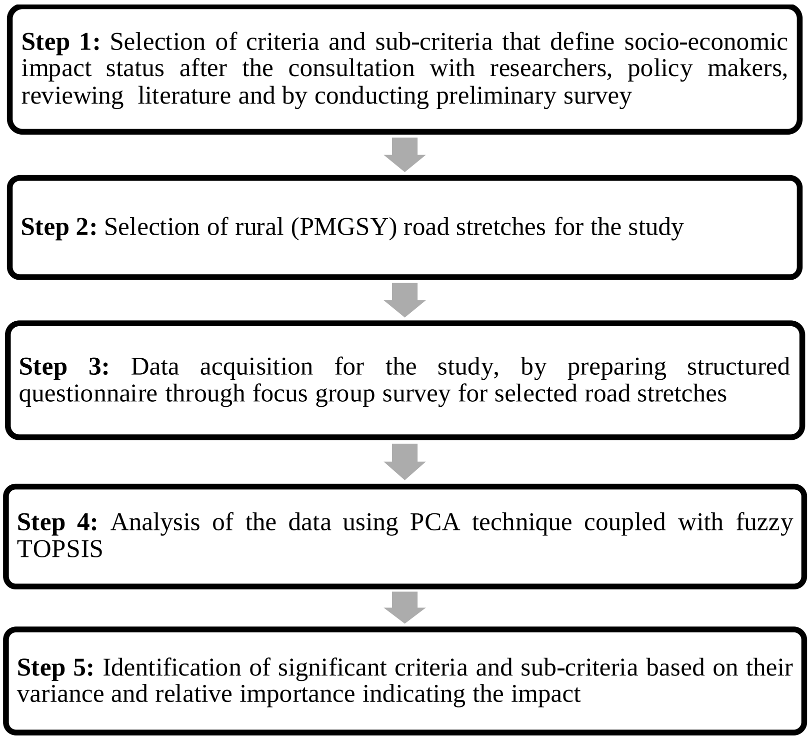

Fig. 3.3 Methodology adopted for the study

3.3.2 Principal component analysis

Principal component analysis is a simple eigenvector-based multivariate analysis technique which stipulates anomalies associated with studies incorporating several variables. It explains the internal structure of data by revealing the variance in the dataset. It identifies inputs of each variable to the components (factor loadings) for a given set of data. The main objective is to have optimum linear combination, where primary criteria explain variability in the data set. Mathematically, PCA technique creates components where every component is a linear weighted combination of primary criteria as given in equation (3.1):

$$\begin{aligned} PC_1 &= X_1 \times a_1 + X_2 \times a_2 + \cdots + X_n \times a_n \\ PC_m &= X_{m1} \times a_{m1} + X_{m2} \times a_{m2} + \cdots + X_{mn} \times a_{mn} \end{aligned} \tag{3.1}$$

where X_{m1} represents the amplitude of m^{th} principal component of the n^{th} criteria. The factor-scores from the model are recovered by modifying the structure inferred by equation (3.1) and yields a set of measures for every m principal components: (n = 1, ..., N). The eigenvalue analogues to eigenvector represent variance (v) for every principal component. The components are arranged according to their variance such that the first principal component (PC_1) elucidates the maximal possible extent of variation. It considers the limitation that the summation of the squared amplitudes equals one (i.e., $X_{12}+ X_{22}+...+ X_{n2} = 1$). The amount of variance accounted by every single component to the total variation in the original data set is given as i/n. It is the summation of the eigenvalues and is equivalent to the total number of criteria in the original dataset.

The methodology of PCA is explained through two steps. In this study, PCA analysis has been performed using SPSS (statistical package for the social sciences) and mono plots have been drawn using "Analyse it-2016' software. Step-wise process of PCA is as explained below.

<u>Step 1:</u> Data Processing

The data points corresponding to 33 sub-criteria have been gathered through focus-group discussions for every selected habitation.

- These collected data points are used as input for principal components analysis (PCA). Before the analysis is performed, the data gathered through focus group panel is made consistent enough on a scale of 0 to 1.
- Further, in employing SPSS, the eigenvalues and the corresponding eigenvectors for the data set are evaluated, which are helpful in determining the variance in data set caused by criteria in terms of principal components.
- The cumulative percent of variance and their respective eigenvalues with respect to principal components (PCs) are shown below in Table 3.2.
- It is observed that cumulative percent variance (88.85%) is contributed by first 10 components. Fig. 3.4 shows the scree plot to demonstrate the variance and cumulative variance with respect to principal components.

Table 3.2 Percentage variability and cumulative variability by the components

Components observing the SEIA sub-criteria	Eigenvalues	% of Variance	Cumulative %
1	5.020	15.213	15.213
2	4.756	14.412	29.625
3	3.924	11.889	41.514
4	3.535	10.712	52.226
5	2.670	8.091	60.317
6	2.538	7.690	68.007
7	2.212	6.702	74.709
8	1.907	5.778	80.487
9	1.507	4.567	85.055
10	1.252	3.794	88.849

Step 2: PCA Interpretation

PCA outputs are tabulated as factor-scores or in the form of sub-criteria weights. Component loading measures the extent of proximity between principal components and sub-criteria, largest the loading either positive or negative represents the significance of the component. Positive loading depicts that the input of sub-criteria augments with the increase in loading of the component, and negative represents reduction. Moreover, sub-criteria with positive loading/weight signifies higher score whereas that with negative value represents a lower score. The study retains 10 PCs which are responsible for a total variance of 88.85%.

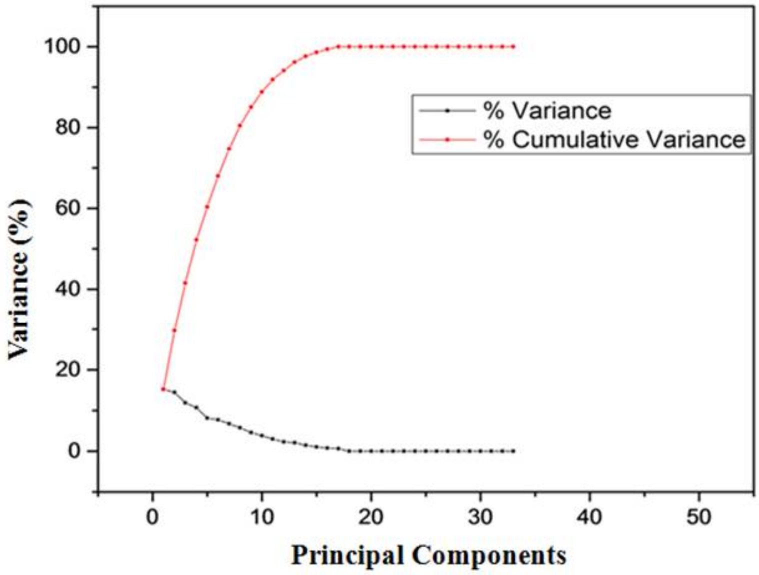

Fig. 3.4 Scree plot depicting the proportion of variance Vs principal components

The scree plot (Fig. 3.4) assists in identifying relevant components which are to be retained for further analysis. All components with eigenvalues greater than one are generally retained. To support this, Monte Carlo PCA tool has also been applied parallelly to perform the analysis. The tool calculates eigenvalues within the specified boundary condition, which are to be compared with eigenvalues obtained from PCA.

Rotated component matrix (Table 3.3) is the key output obtained from PCA; it exhibits the correlation score of different sub-criteria with respect to retained components. In the current study, it is observed that the first component accounts 15.213% of the total variance and is contributed by higher loadings of sub-criteria such as agriculture self-employed (I_{SA}), frequency of public transportation (T_{PUBF}), percent of male (E_{MAS}) and female (E_{FAS}) children attending school. Similarly, the remaining 9 PCs constitute about 73.64% of the total variability. The loading score of each sub-criterion with respect to a given component is illustrated in Table 3.3.

Table 3.3 Rotated component matrix

	Components									
Sub-criteria	1	2	3	4	5	6	7	8	9	10
T_{PUBF}	-0.861	-	-	-	-	-	-	-	-	-0.332
I_{SA}	0.849	-	-	-	-	-	-	-	-	-
E_{MAS}	0.811	-	0.361	-	-	-	-	-	-	-
E_{FAS}	0.788	-	-	-	-	-	-0.327	-	-	-
T_{PVTTC}	0.593	-	-0.363	-	-	-0.558	-	-	-	-
T_{TTPUB}	-0.501	-	0.328	0.402	-	-	-0.336	-	0.480	-
T_{PUBTU}	-	0.903	-	-	-	-	-	-	-	-
T_{PUBTC}	-	0.806	-	-	-	-	-	-	-	-
H_{HAM}	-	-0.778	-	-	-	-	-	-	-	-
E_{FLR}	-	-	0.912	-	-	-	-	-	-	-
I_{WA}	-	-	0.819	-	-	-	-	-	-0.330	-
E_{AM}	-	-	-0.583	-0.560	-	-	0.411	-	-	-
I_{II}	-	-	-	0.936	-	-	-	-	-	-
H_{CA}	-	-	-	0.739	-	-	-	-	-	-
Q_{TVO}	-	-	-	-	-0.792	-	-	-	0.473	-
E_{TT}	-	-	-0.374	-	-0.728	-	-	-	-	-
I_{UA}	-	-	-	-	-0.723	-	-	-	-	-
Q_{SGIV}	-	-	-	-	0.551	0.515	-	-	-	-
Q_{PPO}	0.491	-	-	-	0.532	-	-	-0.416	-	-
Q_{L}	-	-	-	-	-	0.812	-	-	-	-
Q_{SGOV}	-	-	-	-	-	0.751	-	0.397	-	-
E_{PRISA}	-	-	-	-	-	-	0.874	-	-	-
I_{UNA}	-	-	-	0.508	-	-	-0.613	-	-	-
E_{PRESA}	-	-	0.482	-	-	0.333	0.599	-	-	-
E_{MLR}	-	-	-	-	-	-	-	0.851	-	-
H_{FU}	-	-	-	0.331	-	-	-	0.818	-	-
H_{TT}	-	-0.362	-	-	-	-0.437	-	-0.449	-	-
H_{HSANT}	-	-0.446	-	-	-	-	-	-	0.732	-

I_{HH}	-	-	-	-0.339	-	-	-	-	0.712	-
I_L	-	-	0.459	0.360	-	-	-	-	-0.468	0.446
T_{TTPVT}	-	-0.489	-	-	-	-	-	-	-	0.749
H_{PHCA}	-	-0.317	-	-	0.485	-0.306	-	-	-	0.605
T_{PVTTU}	0.442	-	0.308	-	-	-0.316	-	-	-	0.585

A 2-dimensional monoplot shown below in Fig. 3.5 is a representation of component loadings as coefficients of the two principal components. It assists in visualizing the interrelationships among the sub-criteria. Positive correlation is indicated when the vectors representing the sub-criteria pointing away from the origin of the monoplot are in the same direction. Negative correlation is observed when they are at 180°angle to each other. From the first quadrant of the monoplot of Fig. 3.5, the criteria, viz., access to mode of transport for health facility (H_{HAM}), travel time to reach health centres (H_{TT}), availability of primary health care center (H_{PHCA}) have positive correlation with each other. The monoplot also represents type of correlation among sub-criteria.

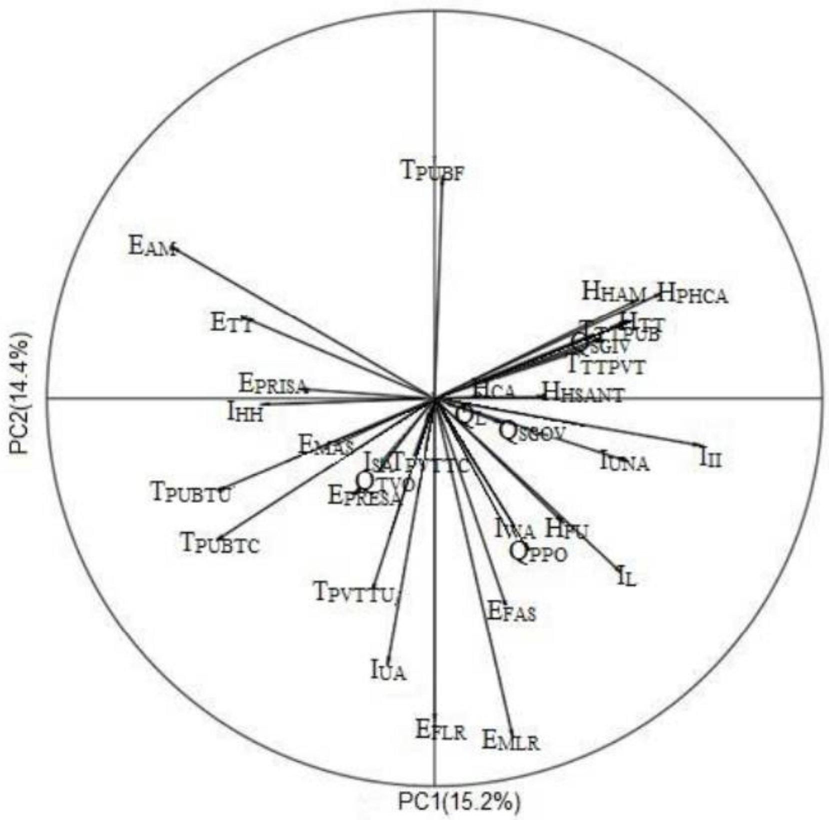

Fig. 3.5 Monoplot representing the correlation

Fig. 3.6 shows loading pattern in 2-dimensional space. It is the component plot in rotated space depicting how the sub-criteria are closely related and their relationship with PCs.

3.3.3 Fuzzy-TOPSIS for ranking the indicators

The study employs PCA for identifying key indicators based on their variance. However, inter-correlation between sub-criteria is of indistinct in nature. This necessitates identifying key sub-criteria by accounting for their relative importance (ranks) and is achieved by employing analytical techniques. Cross-correlation, step-wise approach, multi-criteria techniques like analytical hierarchy process (AHP), fuzzy AHP and fuzzy-TOPSIS are some of the well-known analytical techniques.

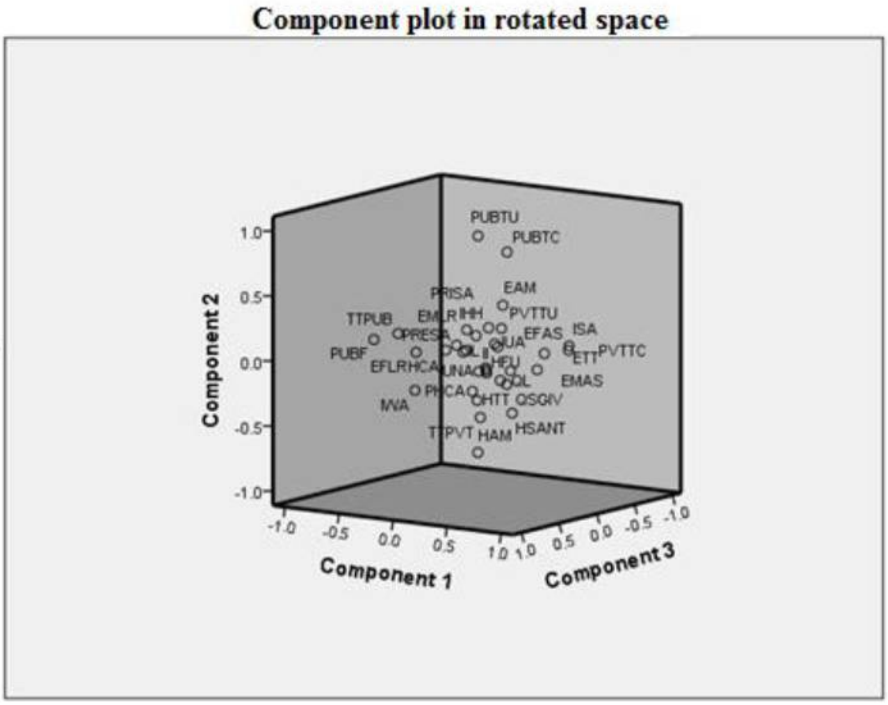

Fig. 3.6 Loading pattern in 2-dimensional space

However, this study applies fuzzy-TOPSIS approach to identify the key criteria by ranking them. It accounts for the change in satisfaction level before and after the deliverance of rural roads. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is one of the recognized ranking techniques among other MCDM methods and has been first considered by (Yoon and

Hwang, 1981). The hypothesis of the TOPSIS methodology is to identify the ideal and nadir solutions (Liang, 1999; Liang and Meng, 2019) and is based on the logic of comparative proximity.

It is observed as the distance of the sub-criteria to the ideal (nadir) point, which are to be ranked based on their priority. In the present study, the collected data is based on human perception and judgments which exhibits fuzziness. To overcome the fuzziness (uncertainty) associated with the data, the concept of fuzzy set theory has been integrated with the TOPSIS technique. This facilitates the assessment of socio-economic impacts (SEI) in a comprehensive way. The methodology followed for the study is shown in Fig. 3.7.

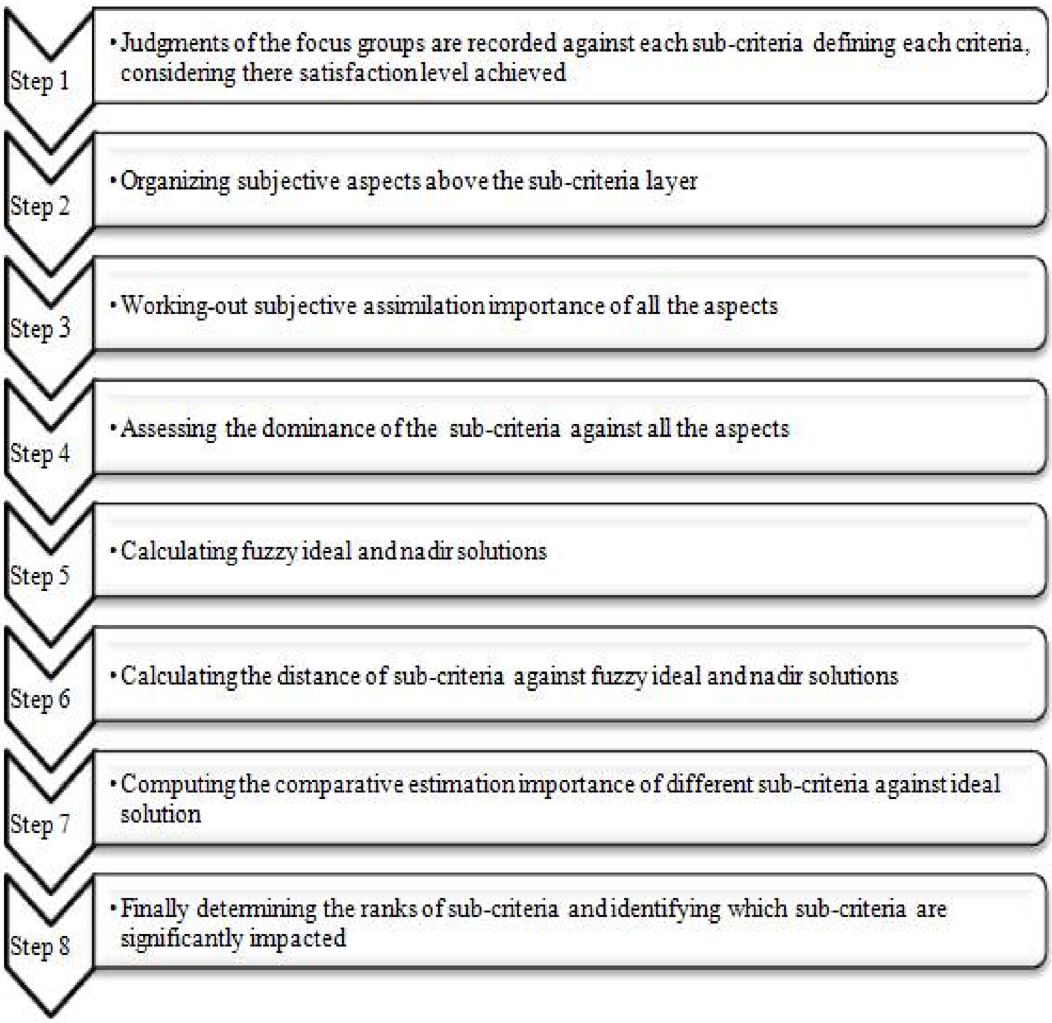

Fig. 3.7 Methodical procedure followed in Fuzzy-TOPSIS

The data process of the fuzzy-TOPSIS approach applied for the study herein as follows:

Step 1: The first step is to identify the important sub-criteria, defining criteria employed in SEIA. They are assessed based on the satisfaction level achieved by the rural inhabitants. The study considers two aspects, viz., satisfaction level 1(SL_1) (before) and satisfaction level 2 (SL_2) (after). The aspects are of positive in nature and represent the satisfaction level of inhabitants before and after the construction of PMGSY roads.

Step 2: Next step is to calculate the comparative importance weights of the aspects. A pair-wise comparison matrix is developed. The study employs a fuzzy analytic hierarchy process (FAHP) approach to acquire the integrated weights for further assessment based on the linguistic ratings from the experts. Below is the pair-wise comparison matrix for two aspects considered, as shown below in Table 3.4.

Table 3.4 Fuzzy pair-wise comparison matrix of aspects

	SL_1			SL_2		
SL_1	1.00	1.00	2.00	0.20	0.25	0.33
SL_2	4.00	5.00	6.00	1.00	1.00	2.00

The geometric mean fuzzy comparison value for the SL_1 aspect obtained is ($\widetilde{r_1}$ = 1.20, 1.30, 2.30). Further, the fuzzy weight ($\widetilde{w_2}$ = 0.12, 0.17, 0.38) for this aspect is obtained using equation (3.2). Finally, the crisp weight for the aspect is obtained using geometric mean integration representation approach (GMIR) (Chen and Hsieh, 2000) as 0.20. Fuzzy weights of the aspects obtained for the study are shown below in Table 3.5. Further, for a given triangular fuzzy number (TFN), the defuzzified weight is calculated by employing GMIR approach by employing equation (3.3). The calculated weights are normalized, which are given below in Table 3.6.

$$\widetilde{w_i} = \sum_1^m \tilde{a}_{ij} \otimes [\sum_{q=1}^m \sum_{r=1}^m \tilde{a}_{qr}]^{-1}, \ i = 1, \ldots, m. \tag{3.2}$$

$$r(a) = \frac{a_1 + 4a_2 + a_3}{6} \tag{3.3}$$

where $\tilde{a} = (a_1, a_2, a_3)$ be the triangular fuzzy number.

Table 3.5 Fuzzy weights for aspect SL_1 and SL_2

Fuzzy weights			
SL_1	0.12	0.17	0.38
SL_2	0.48	0.83	1.29

Table 3.6 Defuzzified and normalized weights of aspects

	SL_1	SL_2
Defuzzified weights	0.20	0.85
Normalized Weight	0.18	0.81

Step 3: The next step in the process is to evaluate the dominance of the sub-criteria in relation with respective aspects. These are acquired by using perception ratings mentioned in linguistic terms, by the focus groups. The mean dominance rating for every sub-criterion is assessed by employing arithmetic mean approach. Table 3.7 shows the dominance of sub-criteria with respective aspects SL_1 and SL_2.

Let, $D_{ic}^{f} = (C_{ic}^{f}, A_{ic}^{f}, B_{ic}^{f})$ is the fuzzy dominance rating of the i^{th} sub-criteria in reference to the c^{th} subjective aspect, evaluated by the f^{th} focus group, where (i = 1, …., m; c = 1, …, p; f = 1, …, n). The mean fuzzy dominance rating of the i^{th} sub-criteria with reference to the c^{th} subjective aspect evaluated by the f^{th} focus-group is evaluated as given below:

$$\left[\frac{\sum_{f=1}^{n} C_{ic}^{f}}{n}, \frac{\sum_{f=1}^{n} A_{ic}^{f}}{n}, \frac{\sum_{f=1}^{n} B_{ic}^{f}}{n}\right]$$

Table 3.7 Dominance of sub-criteria V/s aspects (SL_1) and (SL_2)

Sub-criteria	Fuzzy dominance ratings (SL_1)			Fuzzy dominance ratings (SL_2)		
	C	A	B	C	A	B
T_{TTPUB}	1.37	2.26	3.26	2.16	3.16	4.16
T_{TTPVT}	1.37	2.16	3.21	2.00	2.95	3.95
T_{PUBTU}	1.53	2.37	3.37	2.32	3.32	4.32

T_{PVTTU}	1.42	2.21	3.21	1.95	2.89	3.95
T_{PUBF}	1.47	2.32	3.32	2.05	3.05	4.05
T_{PUBTC}	1.63	2.58	3.58	2.74	3.74	4.74
T_{PVTTC}	1.32	2.21	3.21	2.11	3.05	4.05
I_{II}	1.79	2.74	3.74	2.37	3.37	4.37
I_{HH}	1.58	2.47	3.47	2.21	3.11	4.21
I_{SA}	1.95	2.95	3.95	2.84	3.84	4.84
I_{WA}	1.16	1.95	2.95	1.84	2.84	3.84
I_{L}	1.16	1.84	2.84	1.42	2.32	3.32
I_{UA}	1.42	2.37	3.37	2.21	3.21	4.21
I_{UNA}	1.32	2.05	3.05	1.63	2.42	3.47
H_{FU}	2.42	3.42	4.42	3.63	4.63	5.63
H_{CA}	1.21	1.58	2.58	1.47	2.37	3.37
H_{PHCA}	1.16	1.53	2.53	1.32	2.21	3.21
H_{HAM}	1.26	2.11	3.11	2.00	3.00	4.00
H_{TT}	1.16	1.95	2.95	1.84	2.74	3.74
H_{HSANT}	1.21	1.95	2.95	1.68	2.68	3.68
E_{MLR}	1.42	2.32	3.32	2.00	3.00	4.00
E_{FLR}	2.26	3.26	4.26	2.84	3.84	4.84
E_{MAS}	2.37	3.37	4.37	3.37	4.37	5.37
E_{FAS}	1.79	2.63	3.63	2.89	3.68	4.89
E_{AM}	1.42	2.26	3.26	2.00	3.00	4.00
E_{TT}	1.47	2.37	3.37	2.16	3.11	4.11
E_{PRESA}	1.47	2.32	3.32	2.16	3.00	4.00
E_{PRISA}	1.37	2.16	3.16	1.74	2.63	3.63
Q_{PPO}	1.68	2.68	3.68	3.00	4.00	5.00
Q_{TVO}	1.84	2.84	3.84	3.16	4.16	5.16
Q_{L}	1.74	2.74	3.74	2.68	3.68	4.68
Q_{SGIV}	1.37	2.32	3.32	2.00	3.00	4.00
Q_{SGOV}	1.42	2.32	3.32	2.37	3.37	4.37

Step 4: In this step, ideal and nadir solutions are computed. The ideal and nadir solutions are established on the hypothesis of comparative proximity. They are observed as the distance of sub-criteria *i* to the ideal (nadir) solutions and are ranked accordingly (Liang, 1999; Liang and Meng, 2019). As all the sub-criteria are positive in nature, the standardized fuzzy dominance rating D_{ij} (max) of the i^{th} sub-criteria with respect to aspect *j* is evaluated as shown in equation (3.4), where $\Delta_j = \max(B_{ij})$.

$$D_{ij} = \left(l_{ij}, m_{ij}, k_{ij}\right) = \left[\frac{C_{ij}}{\Delta_j}, \frac{A_{ij}}{\Delta_j}, \frac{B_{ij}}{\Delta_j}\right] \tag{3.4}$$

Table 3.8 illustrates the standardize dominance ratings of sub-criteria obtained with respect to aspects SL_1 and SL_2.

Table 3.8 Standardize dominance rating of sub-criteria V/s aspects (SL_1) and (SL_2)

Sub-criteria	Fuzzy dominance ratings (SL_1)			Fuzzy dominance ratings (SL_2)		
	C	A	B	C	A	B
T_{TTPUB}	0.31	0.51	0.74	0.38	0.56	0.74
T_{TTPVT}	0.31	0.49	0.73	0.36	0.52	0.70
T_{PUBTU}	0.35	0.54	0.76	0.41	0.59	0.77
T_{PVTTU}	0.32	0.50	0.73	0.35	0.51	0.70
T_{PUBF}	0.33	0.52	0.75	0.36	0.54	0.72
T_{PUBTC}	0.37	0.58	0.81	0.49	0.66	0.84
T_{PVTTC}	0.30	0.50	0.73	0.37	0.54	0.72
I_{II}	0.40	0.62	0.85	0.42	0.60	0.78
I_{HH}	0.36	0.56	0.79	0.39	0.55	0.75
I_{SA}	0.44	0.67	0.89	0.50	0.68	0.86
I_{WA}	0.26	0.44	0.67	0.33	0.50	0.68
I_{L}	0.26	0.42	0.64	0.25	0.41	0.59
I_{UA}	0.32	0.54	0.76	0.39	0.57	0.75
I_{UNA}	0.30	0.46	0.69	0.29	0.43	0.62
H_{FU}	0.55	0.77	1.00	0.64	0.82	1.00

H_{CA}	0.27	0.36	0.58	0.26	0.42	0.60
H_{PHCA}	0.26	0.35	0.57	0.23	0.39	0.57
H_{HAM}	0.29	0.48	0.70	0.36	0.53	0.71
H_{TT}	0.26	0.44	0.67	0.33	0.49	0.66
H_{HSANT}	0.27	0.44	0.67	0.30	0.48	0.65
E_{MLR}	0.32	0.52	0.75	0.36	0.53	0.71
E_{FLR}	0.51	0.74	0.96	0.50	0.68	0.86
E_{MAS}	0.54	0.76	0.99	0.60	0.78	0.95
E_{FAS}	0.40	0.60	0.82	0.51	0.65	0.87
E_{AM}	0.32	0.51	0.74	0.36	0.53	0.71
E_{TT}	0.33	0.54	0.76	0.38	0.55	0.73
E_{PRESA}	0.33	0.52	0.75	0.38	0.53	0.71
E_{PRISA}	0.31	0.49	0.71	0.31	0.47	0.64
Q_{PPO}	0.38	0.61	0.83	0.53	0.71	0.89
Q_{TVO}	0.42	0.64	0.87	0.56	0.74	0.92
Q_{L}	0.39	0.62	0.85	0.48	0.65	0.83
Q_{SGIV}	0.31	0.52	0.75	0.36	0.53	0.71
Q_{SGOV}	0.32	0.52	0.75	0.42	0.60	0.78

Further, the fuzzy ideal and nadir solutions for sub-criteria with respect to aspects (SL_1) and (SL_2) are computed with respect to representation values r (D_{ij}), by employing GMIR approach. The fuzzy ideal and nadir solutions are defined as; fuzzy ideal solution (I) = ($D_1^+, D_2^+, \ldots, D_j^+, \ldots, D_c^+$) and nadir solution as (N) = ($D_1^-, D_2^-, \ldots, D_j^-, \ldots, D_c^-$). Table 3.9 shows the fuzzy ideal and nadir ratings obtained.

Table 3.9 Fuzzy ideal and nadir values for sub-criteria with respect to aspects (SL_1 and SL_2)

Aspects	Fuzzy Ideal values			Fuzzy Nadir values		
SL_1	0.55	0.77	1.00	0.23	0.23	0.45
SL_2	0.65	0.83	1.00	0.17	0.17	0.35

Step 5: Evaluating sub-criteria distances, with reference to fuzzy ideal and nadir solutions using equation (3.5) and equation (3.6). Table 3.10 depicts distance of sub-criteria with reference to

fuzzy ideal and nadir solutions.

$$d_i^+ = \sqrt{\sum_{j=1}^{c}\left[\left(\beta_j\right)^2 \times \left(\alpha_M \times \left(D_j^+, D_{ij}\right)\right)^2\right]} \tag{3.5}$$

$$d_i^- = \sqrt{\sum_{j=1}^{c}\left[\left(\beta_j\right)^2 \times \left(\alpha_M \times \left(D_j^-, D_{ij}\right)\right)^2\right]} \tag{3.6}$$

where $i = 1, 2, \ldots, k$.

Step 6: Final step is to assess the rank of sub-criteria with respect to comparative estimate in relation to an ideal solution (for the present study). The comparative estimate (CE^*) in relation to the ideal solution is assessed by employing equation (3.7):

$$CE_i^* = \frac{d_i^-}{d_i^- + d_i^+} \tag{3.7}$$

Table 3.10 Distance of sub-criteria V/s fuzzy ideal and nadir Solutions

Sub-criteria	d_i^-	d_i^+	Sub-criteria	d_i^-	d_i^+
T_{TTPUB}	0.540	0.540	H_{HAM}	0.413	0.543
T_{TTPVT}	0.544	0.544	H_{TT}	0.398	0.548
T_{PUBTU}	0.537	0.537	H_{HSANT}	0.394	0.551
T_{PVTTU}	0.544	0.544	E_{MLR}	0.414	0.543
T_{PUBF}	0.542	0.542	E_{FLR}	0.394	0.532
T_{PUBTC}	0.532	0.532	E_{MAS}	0.515	0.531
T_{PVTTC}	0.542	0.542	E_{FAS}	0.468	0.529
I_{II}	0.536	0.536	E_{AM}	0.475	0.543
I_{HH}	0.538	0.538	E_{TT}	0.421	0.540
I_{SA}	0.532	0.532	E_{PRESA}	0.416	0.541
I_{WA}	0.547	0.547	E_{PRISA}	0.392	0.551
I_L	0.560	0.560	Q_{PPO}	0.484	0.531

I_{UA}	0.539	0.539	Q_{TVO}	0.496	0.531
I_{UNA}	0.554	0.554	Q_L	0.461	0.533
H_{FU}	0.533	0.533	Q_{SGIV}	0.414	0.543
H_{CA}	0.559	0.559	Q_{SGOV}	0.438	0.537
H_{PHCA}	0.564	0.564			

Table 3.11 Rank of sub-criteria with respect to comparative estimate (*CE**)

Sub-criteria	CE^*	Rank	Sub-criteria	CE^*	Rank
T_{TTPUB}	0.4396	15	H_{HAM}	0.4319	22
T_{TTPVT}	0.4303	24	H_{TT}	0.4204	27
T_{PUBTU}	0.4471	12	H_{HSANT}	0.4168	28
T_{PVTTU}	0.4287	25	E_{MLR}	0.4325	20
T_{PUBF}	0.4349	18	E_{FLR}	0.4719	5
T_{PUBTC}	0.4658	8	E_{MAS}	0.4921	2
T_{PVTTC}	0.4353	17	E_{FAS}	0.4832	3
I_{II}	0.4504	10	E_{AM}	0.4324	23
I_{HH}	0.4405	14	E_{TT}	0.4381	16
I_{SA}	0.4711	6	E_{PRESA}	0.4345	19
I_{WA}	0.4241	26	E_{PRISA}	0.4161	29
I_L	0.4004	32	Q_{PPO}	0.4767	4
I_{UA}	0.4423	13	Q_{TVO}	0.4693	7
I_{UNA}	0.4077	30	Q_L	0.4640	9
H_{FU}	0.5011	1	Q_{SGIV}	0.4325	21
H_{CA}	0.4022	31	Q_{SGOV}	0.4492	11
H_{PHCA}	0.3947	33			

The study evaluates the most substantial sub-criteria and ranks them based on their relative importance. Table 3.11 represents the ranks obtained by the sub-criteria based on the perceptions of the focus groups for the selected habitations. It has been observed from the above evaluation that the sub-criteria contributing to health facility (e.g. use of health facility (H_{FU})), and sub-criteria contributing to education facility (percent of female children attending school (E_{FAS}), literacy rate of female (E_{FLR}), literacy rate of male (E_{MLR})) are ranked the most significant. This depicts there has been significant change in these sub-criteria. These are followed by the sub-criteria contributing to the quality of neighbourhood (social environment) (i. e., Q_{TVO}, Q_L, and Q_{SGOV}) and Transport facility (i.e., T_{TTPUB}, T_{PUBTC}, and T_{PVTTC}). Furthermore, from the analysis, it has also been inferred that the sub-criteria contributing to a health facility (i.e., availability of clinic (H_{CA}), availability of primary health Centre (P_{HCA}), travel time to reach health facility (H_{TT})) have gained lower ranks, which depicts inhabitant's dissatisfaction towards these sub-criteria.

3.4 Result and discussion

This study presents a novel mixed-method approach which integrates PCA with fuzzy-TOPSIS technique to assess the performance of socio-economic indicators/criteria after the deliverance of rural roads. PCA considers quantitative data whereas fuzzy-TOPSIS accounts qualitative data. The study first explores the socio-economic indicators in terms of their variance which is then followed by ascertaining them based on their relative importance. From PCA assessment, it has been observed that the first principal component (PC) accounts 15.21% variance which is contributed by higher loadings of sub-criteria, viz., the frequency of public transportation (T_{PUBF}), self-employed from agriculture (I_{SA}), and percent of male children attending school (E_{MAS}). This depicts that after the construction of PMGSY roads there has been a substantial change in these sub-criteria. The higher loading of sub-criterion I_{SA} represents that there is a possible increase in dependency of rural inhabitants on agriculture due to availability of resources as well as enhanced physical access to the nearest markets. It can also be put forth that it may be due to decreased transportation and production cost of agricultural produce.

Moreover, higher loading of sub-criteria E_{MAS} followed by E_{FAS} depicts increased accessibility to the schools along with a considerable reduction in travel time to reach them. As travel time plays a significant role in case female students, which increases their possibility to attend school. Similarly, positive changes such as an increase in number of public transportation units (T_{PUBTU}) along with a midcore change in the cost of travel incurred by public transportation (T_{PUBTC}) are observed. This is evident from their loadings contributing to the second PC. Consequently, it is also observed that the sub-criteria, viz., Q_{SGOV}, Q_L, Q_{TVO} contributing to the quality of neighbourhood (social environment) indicator show higher loadings. This depicts that there has been a substantial change in the living and social condition of inhabitants. Moreover, the loadings of these sub-criteria also represent that possible positive change in the livibility conditions of inhabitants within the community and enabling them to involve in social gatherings. It also indicates positive change in the quality of life of marginal groups (especially women).

However, along the positives changes, the analysis also depicts no change in the condition of some sub-criteria (e.g., H_{TT}), which is evident from their lower loadings (Table 3.3). The lower loadings in case of sub-criteria travel time to reach health facility are probably because no subsequent change in the travel time to reach health facility available to the inhabitants even after road construction. The possible reason is that the inhabitants are trying to avail proper treatment which needs a longer distance to be commuted. However, to have enhanced comprehension in supplementary to PCA about the status of the sub-criteria, the sub-criteria are ranked based on the comparative satisfaction level of rural inhabitants. Fuzzy-TOPSIS approach is employed to elucidate the most significant one. It is observed that the sub-criteria contributing to the quality of neighbourhood (social environment), viz., Q_L, Q_{SGOV}, Q_{TVO} have been ranked with high priority. This depicts positive change in living standard of inhabitants. Positive change can also be inferred in case of education facility indicator based on higher ranks of sub-criteria, viz., E_{FAS}, E_{FLR}, and E_{MLR}. This is because possible reduction in travel time and the ability of female students to avail education facility with ease, which is also apparent from PCA analysis.

Furthermore, it is also understood that some of the sub-criteria, viz., T_{TTPUB}, T_{PUBF}, T_{PUBTC} contributing to transport facility indicator have been ranked as mediocre. A similar pattern is also observed in case of the sub-criteria contributing to health facility criteria (e.g., H_{FU}) as well as income status criteria (i.e., I_{HH}, I_{UA}). The sub-criteria viz., H_{TT}, H_{CA}, H_{PHCA} contributing to health facility criteria have been ranked low. This is also observed in the case of sub-criteria (i.e., I_L, I_{UNA}, I_{WA}) contributing to income status indicator. This depicts little to no change in the status of these sub-criteria which substantiate our analysis done using PCA.

3.5 Summary

Assessment of socio-economic impacts instigated by the deliverance rural roads is of prime importance from the viewpoint of sustainable rural development. It reveals necessary knowledge about the potential socio-economic and cultural impacts on the lives of rural habitants and their communities. It assists concerned decision makers in finding the ways to mitigate or prevent adverse or insignificant impacts from happening. Moreover, it also emphasizes on maximizing beneficial impacts, achieved by the provision of the planned forum. SEIA involves several criteria (i.e., qualitative and quantitative) and their interdependencies. Thus, ascertaining and exploring them creates a need for a systematic tool so that comprehensive assessment can be achieved at the regional level. Although, literature suggests several techniques (experimental and quasi-experimental), yet they lag to accommodate the problem of biases arising from real life data.

Considering this, the present study proposes a novel mixed-method approach which integrates multivariate analysis with fuzzy MCDM technique. Here, PCA considers quantitative data whereas fuzzy-TOPSIS accounts qualitative data. The proposed approach accommodates the advantages of mixed-method design like its ability to attain any kind of changes according to the necessity of the study to be conducted. As value addition, it increases the reliability of SEIA methodology. It deepens the understanding of SEIs to be perceived by decision-makers and stakeholders with ease. From the analysis viewpoint of the present study, it is revealed that the PMGSY roads have contributed significantly in the upliftment of rural life. They instigated economic growth along with a change in the livibility condition of the rural population which can be well-perceived form the relative importance and variance of sub-criteria contributing to income status and quality of neighbourhood criteria. Furthermore, the analysis also points out that the concerned decision makers are required to take necessary initiatives in promoting non-farm activities to promote livelihood diversification and making rural population self-sustainable. Moreover, from the analysis, it is also observed that proper distribution of health and education facility available to rural inhabitants is needed, which is important from the viewpoint of overall rural development.