

An Efficient Feature Selection Technique using Genetic Algorithm for Activity Recognition of Elder People

Shreya Gupta

B.Tech Scholar,
Department of Computer Engineering & Technology
Guru Nanak Dev University, Amritsar

ABSTRACT

Activity recognition technique is critical for the elderly people. To avoid the need of full time care giving service, the actual trend is to encourage elderly to stay living autonomously in their homes as long as possible. Reliable activity recognition techniques can improve life safety of the elderly and boost their confidence by immediately alerting fall cases to caregivers. Machine learning techniques are extensively used to predict the activity recognition of elder people. This paper has proposed the integration of the feature selection technique and genetic algorithms to improve the accuracy rate for detection of activity recognition of elder people as well as substantial features of activity recognition of elder people. The experimental results brings about the proposed technique that clearly shown the fact that proposed technique outperforms over the existing methods.

Keywords

Machine Learning, Water mining, Activity recognition of elder people, Feature Selection and Genetic Algorithm.

INTRODUCTION

Use of machine learning is increasing in activity recognition of elder people area in these days. As we know various applications of activity recognition of elder people machine learning are gaining popularity day by day [1]. Mainly machine learning concept is basically used in activity recognition of elder people management and classification analysis to improve the success rate. Main problem faced by the activity recognition of elder people sector is to improve the performance of machine learning algorithms by using valuable technique. Activity recognition of elder people falls under respective classes [2].

Machine learning helps its users deduce important information from huge databases. In medical stream, practitioners make use of huge patient data. Any effective medical treatment is achieved after complete survey of ample amount of patient data. But practitioners usually faced with the obstacle of deducing pertinent information and finding certain trend or pattern that may further help them in the analysis or treatment of any disease [3]. Machine learning is such a tool which sifts through that voluminous data and presents the data of essential nature. In this paper, we have designed a five-step machine learning model that will help medical practitioners on determining the appropriate drug to be used in ministration for epilepsy [4].

Activity recognition techniques play significant role in daily life of elderly people [5]. To avoid the need of full time care

giving service, the actual trend is to encourage elderly to stay living autonomously in their homes as long as possible. Reliable activity recognition techniques can improve life safety of the elderly and boost their confidence by immediately alerting fall cases to caregivers [6].

1.1 Machine Learning in Support Vector Machines

Machine learning is the subfield of computer science that offers computers the capability to learn without having to be explicitly programmed. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms is unfeasible. For example spam filtering, optical character recognition (OCR) search engines and computer vision. In support vector machines, data point is viewed as p-dimensional vector (a list of p numbers). Consider there are some data points and they each belong to one of the two classes and we want to know that to which class the new data point will belong to and we also want to know that these data points can be separated by (p-1) hyper planes. This is called as linear classifier. There are many hyper planes that might classify the data. The best hyper plane is the one which provides the largest separation between the two classes. The hyper plane that is having the maximum distance from it to nearest data point on each side is chosen. The hyper plane is known as the maximum-margin hyper plane and the linear classifier is known as a maximum margin classifier.

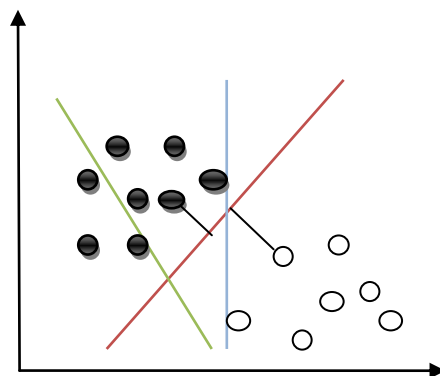


Fig. 1. Depiction of Hyper planes

H₁-It does not separate the classes.

H₂-It separates the classes but with small margin.

H₃-It separates the classes with maximum margin.

1.2 Activity recognition of elder people

Sensor devices are low-power devices equipped with one or more sensors, a processor, memory, a power supply, a radio and an actuator[5]. They have limited energy, computation and short communication range and low bandwidth. Sensor devices have three components, which are sensing, data processing and communicating process. The limited energy is a big problem for WSN and minimising the energy, the energy consumption during network activities is aimed in the studies. Hence, the lifetime of a sensor network can be extended by applying energy efficient protocols[6]. They can sense and collect raw data from the environment and do local processing and computation. After the processing, sensor devices can communicate with each other to perform aggregation. The aggregated data are sent to sinks or base stations. Sinks or base stations make analysis and decisions and link to the outside world via the Internet or satellites[7].

Activity recognition is major problems for aged population representing the fastest growing group of the population. By the year 2025, one-third of the European population will be more than 60 years old[8]. Growing aged population create a financial cost on Public Health Services. Owing to limited resources in Public Health Services, aged people are encouraged to stay living at home instead of living in public care centres. Fall is an obstacle for aged people who are living alone. Children or relatives of these people want to monitor the elder people. Hence, fall detection applications using WSN is one of the studies that aims to monitor the elder people.

Machine learning has attracted a great deal of attention in the information industry and in society as a whole in recent years, because of the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge[9]. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection and customer retention, to production control and science exploration[10].

2. PROPOSED TECHNIQUES

2.1 Feature Selection:

Feature selection, named as variable selection, attribute selection or variable subset selection, is the method of choosing a part of relevant features (variables, predictors) for utilize in design construction. Feature selection attempts to choose the simply sized subset of features depending on the following. It can be:

1. The classification accuracy does not considerably reduce as well as
2. The resulting class allocation, known only the values for the picked features, can be as close as possible to the unique class allocation, known all features.

There are lots of potential great things about variable as well as feature selection: making ease of data visualization as well as data understanding, lowering the measurement and storage requirements, reducing training as well as utilization times, defying the problem of dimensionality to boost recognition performance.

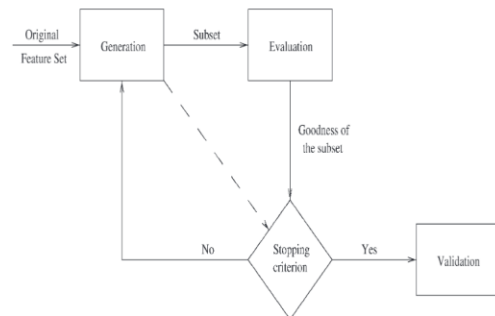


Fig 2: Feature Selection Process

2.2 Genetic Algorithm:

Genetic Algorithms are certainly one of computational models stimulated together with the development and natural Selection processes. They design the most perfect option is together with the problem right data structure called chromosome, genotype or genome signifying the wide ranging solutions, called individuals or creatures or phenotypes. Quite a few genetic operators are provided about bat roosting chromosomes to acquire a quite high optimization together with the problem. An innate algorithm begins employing a arbitrary population of individuals. Up to date inhabitants are changed with the genetic operators' right new generation of individuals. The key element goal of a period may be to keep your *best* individuals, improving the fitness together with the population, until some stopping criteria is achieved. This criterion is usually a fitness threshold, the sheer quantities of generations or lacking improvement.

1. Selection operator's makes using the evaluation function to find out what humans possess the biggest potential. They might persist whilst in the population and used with the other operators.
2. The recombination operators (mutation and crossover) are traditionally utilized to produce new individuals using a few high potential individuals. They are designed to diversify the search process. The widely accepted operators of these kinds are cross-over as well as mutation.
3. The cross-over operator uses 2 if not more fractions of high potential individuals to develop a fresh individual that's appended to your higher generation with all the population.
4. The mutation operator, on contrary, takes one high potential individual as well as really a little alternation in amongst its components. The modern person is usually appended over the next generation with all the population.

3. RELATED WORK

S. Sasikala et.al [1] developed a mimetic algorithm by adding Genetic algorithm and Shapely value for multi-class classification named as SVEGA through managing various dimensional data. A SVEGA model is applied as the feature selection instrument which serves for the improvement of classification engine. Mark E. Borsuk et.al [2] proposed a BOD decay approach which employs Baye's theorem to generate a joint probability distribution for several parameter values contingent on the seen data. Applying this joint distribution, a mathematical integration method is utilized to obtain limited parameter distributions which can be applied to directly evaluating the general plausibility of competitive parameter values. Zexuan Zhu et.al [3] discussed a novel hybrid filter and wrapper feature selection algorithm that

employs filter ranking approach as LS heuristic. The exploratory outcomes shown reveal that the planned technique queries more proficiently and is effective at providing excellent classification accuracy with a couple of characteristics simultaneously. Salisu Yusuf Muhammad et.al [4] developed an appropriate classification model for analyzing as well as classifying activity recognition of elder people in accordance with the machine learning algorithms. Shah Christirani Azhar et.al [5] has discussed three multivariate mathematical practices (CA, PCA, and DA) which aids to identify nine tracking programs situated on the stream into categories of related activity recognition of elder people traits predicated on six picked activity recognition of elder people variables. This purpose possessed a appropriate classification performance of 100%. JanetR. Barclay et.al [6] recommended that the associations between activity recognition of elder people and land cover may potentially increase water body classification techniques, and we suggest improved defenses for these Class Type B watersheds to keep up their AA-like activity recognition of elder people for possible future drinking water supply. ZhenXiang Xing et.al [7] proposed a fuzzy comprehensive evaluation model basically depends on entropy weight method (FCE-EW) that was created to measure the precise state condition of underground activity recognition of elder people. K.Z. Mao et.al [8] proposed a SVM discriminative function pruning analysis (DFPA) algorithm for feature subset selection. The potency of the DFPA strategies have been approved applying two big real life problems. Yaonan Wang et.al [9] developed a middle initialization technique centered on MST to deal with the issue of regional minima. Contrast results display that our initialization approach is preferable to or equivalent with CCIA and kd-tree with regards to sample acceptance rate but computationally costly in certain cases. S. Wechmong khonkon et.al [10] discussed a MLP neural network utilizing the Levenberg-Marquardt algorithm is employed to categorize the activity recognition of elder people of Dust district canals of Bangkok, Thailand. The outcomes show that the neural system accomplishes a higher accuracy classification percentage of 96.52%. Adam Woznica et.al [11] proposed a general framework that focused on various feature designs made from a certain dataset that are able to remove designs from within the models. He found that the closed Apr aggregation method that gives an excellent compromise involving the stability and predictive performances. The all Med strategy that depends on the k-medoids clustering algorithm and hence it is an easy task to implement. Lee Yoot Khuan et.al [12] has shown the back propagation neural network, the modular neural network and the radial basis function network which have been applied to model and thus estimate the Activity recognition of elder people Index. The MNN model was discovered to be the most appropriate model for use to ascertain the WQI, in relation to accuracy and rapidly understanding time. Chamara P Liyanage et.al [13] discussed various classification models for evaluating the activity recognition of elder people to reveal the best accuracy. The BN model has shown the most appropriate accurate accuracy as compared to the MLP model.

4. METHODOLOGY

4.1 Proposed Algorithm

The proposed work includes feature selection technique based genetic algorithm with various machine learning algorithms.

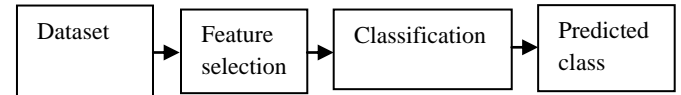


Fig 3: Research Framework

The metrics vectors contain the form: $x_i G = [x_{1,i,G}, x_{2,i,G}, \dots, x_{D,i,G}] = 1, 2, \dots, N$ where G is the generation number.

Step1. Initialization:

1. Initialize upper as well as lower bounds for every metric:

$$x_j^l \leq x_{j,i,1} \leq x_j^u$$

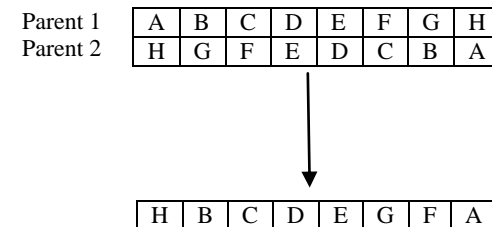
2. Arbitrarily choose the initial metric values consistently on the intervals $[x_j^l, x_j^u]$

Step2. Mutation: Mutation is a process in which a bit involves flipping it, changing 0 to 1 and vice-versa. For example:

Before	0	1	1	1
After	0	1	0	1

1. Every N metric vectors go through mutation, crossover as well as selection.
2. Mutation broad the search space.
3. For a known metric vector x_{iG} arbitrarily choose three vectors $x_{r1,G}$, $x_{r2,G}$, as well as $x_{r3,G}$, for example which indices $i, r1, r2$ and $r3$ are different.
4. Include the weighted variation of two of the vectors to the third $v_{i,G+1} = x_{r1,G} + 1G + (x_{r2,G} - x_{r1,G})$
5. The mutation factor F is a constant from $[0, 2]$.
6. $v_{i,G+1}$ Is called the donor vector.

Step 3: Crossover: Crossover is a process of taking more than one parent solutions and producing a child solution from them. For example:



1. Crossover incorporates achievable solutions from the previous generation.
2. The trial vector u_{iG+1} is constructed from the elements of the target vector, x_{iG} as well as the elements of the donor vector, v_{iG+1}
3. Elements of the donor vector enter the trial vector with probability CR.

$$u_{iG+1} = \begin{cases} v_{iG+1} & \text{if } \text{Rand}_{j_i} \leq CR \text{ or } j = 1_{rand} \\ x_{iG} & \text{if } \text{Rand}_{j_i} > CR \text{ or } j \neq 1_{rand} \end{cases}$$

$$i = 1, 2, \dots, N; j = 1, 2, \dots, D$$

$\text{Rand}_{j_i} \sim U[0,1]$ 1_{rand} is a random integer from $[1, 2, \dots, D]$

1_{rand} ensures that $u_{iG+1} \neq x_{iG}$

Step 4: Selection: Selection is a process that gives preference to improve individuals, permitting them to transfer their genes or individuals to a higher generation. The goodness of every

individual is dependent upon its fitness. Fitness may be based upon goal function or using a subjective judgment.

1. The target vector $x_i G$ is compared with the trial vector $v_i G + 1$ and the one with the lowest function value is admitted to the next generation

$$u_{ji} G + 1 = \begin{cases} u_{ji} G + 1 & \text{iff } (u_{ji} G + 1) \leq x_i G \\ x_i G + 1 & \text{otherwise} \end{cases}$$

2. Mutation, crossover as well as selection continue unless a number of stopping criterion is reached.

In proposed algorithm first it will check what the problem for which it will work is. Next we will collect the data that is required. We have collected activity recognition of elder people dataset from the respective site activity recognition of elder people for activity recognition of elder people classification and for recognition of the activity recognition of elder people class. In next step data selection and transformation is performed. Feature selection technique is helpful for selecting the desired attributes from the list of attributes. Once dataset is loaded we apply the machine learning algorithms such as Naïve Bayes, Decision Table, Random Forest, etc for activity recognition of elder people classification.

5. ANALYSIS OF RESULTS

This section covers the various machine learning algorithms using genetic algorithm between existing and proposed techniques. The projected algorithm is tested on several machine learning algorithms. Some familiar accuracy parameters for activity recognition of elder people have been chosen to show that the efficiency of the projected algorithm is better than the existing techniques. For experimentation and implementation the proposed technique named as genetic algorithm is evaluated using MATLAB tool u2013a. Here we will compare the performance of existing machine learning algorithms with the feature selection technique based on genetic algorithm evaluate the parameters TP rate, FP rate, Precision, Recall and F-measure.

The results are tabulated in the respective table given below. The activity recognition of elder people dataset the features are reduced from the large dataset. The feature selection technique will select only those attributes which are relevant for activity recognition of elder people classification. We have total 30 attributes in the dataset and only 9 attributes are selected for activity recognition of elder people using machine learning process. This reduced dataset is then used for activity recognition of elder people classification where the “numerictonominal” attribute is applied before the classification. The dataset is loaded to classify the activity recognition of elder people and the results are recorded by evaluating the various parameters.

Table 1. Feature selection based Genetic algorithm for activity recognition of elder people

Dataset	No. of attributes	No. of classes
Activity recognition of elder people (without GA)	30	4
Activity recognition of elder people (with GA)	9	4

1. TP Rate:

TPR refers to True Positive Rate and defined as measurement of positive cases which are properly recognized. It is recognition of correctly identified instances. TPR can be expressed by using formula:

$$TP \text{ Rate} = \frac{TP}{TP + FN}$$

2. FP Rate:

FPR is called False Positive Rate and defined as portion of those instances or objects which are imperfectly recognized as positive. FPR can be expressed by using the formula:

$$FP \text{ Rate} = \frac{FP}{FP + TN}$$

3. Precision

Precision is defined as measurement of all positive cases that are identified when making calculations. Higher Precision signifies that an algorithm significantly returned more significant results when compared to irrelevant. Precision can be calculated by using the formula:

$$\text{Precision} = \frac{TP}{FP + TP}$$

4. Recall

Recall is the division of the written documents which are appropriate to the query that have been convincingly recovered. It is also called TP Rate or Sensitivity. It is defined as collection of positive cases. Recall can be expressed as:

$$\text{Recall} = \frac{TP}{TP + FN}$$

5. F-Measure

It contains together precision as well as recall. It is usually utilized to check the accuracy as well as reliability. It computes the mean of precision and recall. Basically, it uses as best and 0 as worst when both precision and recall are used. F-measure can be calculated with using the formula given as:

$$F - \text{Measure} = 2 * \frac{P * R}{P + R}$$

6. ROC Area

It is a just a frequently used graph that summarizes the efficiency of a classifier over-all probable thresholds. It is mainly produced by representing the TP Rate (y-axis) contrary to the FP Rate (x-axis) as it vary the threshold for assigning findings to a certain class. Basically, it uses as best and 0 as worst when both precision and recall are used. It is commonly referred to as Receiver Operating Characteristics.

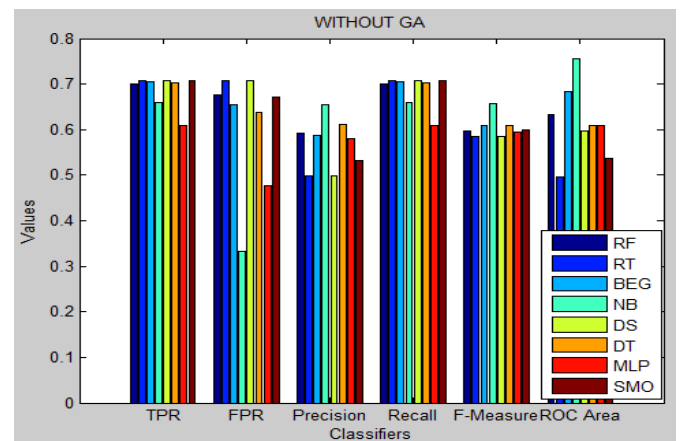


Fig 4: Comparison of TPR, FPR, Precision, Recall, F-Measure and ROC area by each classifier without GA

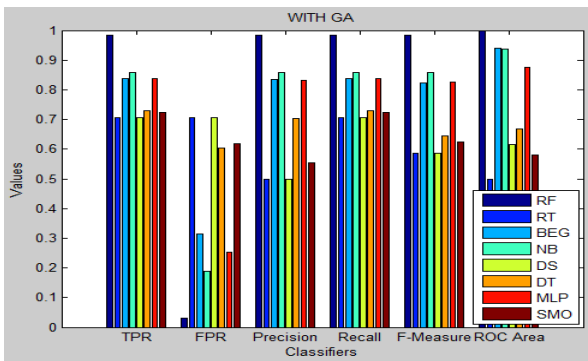


Fig 5: Comparison of TPR, FPR, Precision, Recall, F-Measure and ROC area by each classifier with GA

Fig 5 represents that the random forest with genetic algorithm outperforms because it shows maximum output for classifying the activity recognition of elder people.

6. CONCLUSION

In this paper it represents that existing literature has introduced the detection of elder people activity recognition of elder people has been considered and evaluates by using various state-of-the art algorithms but still there are some issues left to enhance the accuracy rate further for recognition of activity recognition of elder people. So to improve this new method has been proposed i.e. by integrating the feature selection technique and Genetic algorithms that enhances the accuracy rate of activity recognition of elder people as well as substantial features of activity recognition of elder people. The proposed technique has been designed and implemented in weka and Matlab toolbox 2013. The simulation result shows that by applying integration of the feature selection technique and Genetic algorithms by using various parameters i.e. TP Rate, FP Rate, Precision, Recall, F-Measure and ROC Area. This proposed method shows better results as compared to existing results.

REFERENCES

- [1] S. Sasikala, S.A.A. Balamurugana, S. Geetha, "A novel adaptive feature selector for supervised classification", *Information Processing Letters* 117 (2017) 25–34.
- [2] M.E. Borsuk and C.A. Stow, "Bayesian parameter estimation in a mixed-order model of BOD decay," *Water Research*, vol. 34(6), pp.1830-1836, 2000.
- [3] Z. Zhu, Y.S. Ong, M. Dash, Wrapper-filter feature selection algorithm using a memetic framework, *IEEE Trans. Syst. Man Cybern., Part B* 10(4) (2006) 392–404.
- [4] S. Y. Muhammad, M. Makhtar, A. Rozaimie, A. A. Aziz, A. A. Jamal, "Classification Model for Activity recognition of elder people using Machine Learning Techniques", *International Journal of Software Engineering and Its Applications* Vol. 9, No. 6 (2015), pp. 45-52.
- [5] S.C. Azhar, A.Z. Arisa, M.K.Yusoffa, M.F. Ramli, H. Juahir, "Classification of river activity recognition of elder people using multivariate analysis", *International Conference on Environmental Forensics 2015 (iENFORCE2015)*.
- [6] J.R. Barclay, H. Tripp, C.J. Bellucci, G. Warner, A.M. Helton, "Do waterbody classifications predict activity recognition of elder people?", *Journal of Environmental Management* 183 (2016) 1-12.
- [7] Z. Xing, Q. Fu, D. Liu, "Activity recognition of elder people Evaluation by the Fuzzy Comprehensive Evaluation based on EW Method", 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD).
- [8] K.Z. Mao, "Feature subset selection for support vector machines through discriminative function pruning analysis," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 60–67, Feb. 2004.
- [9] Y. Wang, C. Li and Y. Zuo, "A Selection Model for Optimal Fuzzy Clustering Algorithm and Number of Clusters Based on Competitive Comprehensive Fuzzy Evaluation", *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, VOL. 17, NO. 3, JUNE 2009.
- [10] S. Wechmongkhonkon, N. Poomtong, S. Areerachakul, "Application of Artificial Neural Network to Classification Surface Activity recognition of elder people," *International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering* Vol:6, No:9, 2012.
- [11] A. Woznica, P. Nguyen, A. Kalousis, "Model mining for robust feature selection", *KDD '12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and machine learning*, ACM New York, NY, USA, PP 913-921, 2012.
- [12] L. Khuan, N. Hamzah and R. Jailani, "Recognition of Activity recognition of elder people Index(WQI) Based on Artificial Neural Network(ANN)", *Conference on Research and Development Proceedings, Malasia*, 2002, pp. 157-161.
- [13] Singh, Dilbag, Deepak Garg, and Husanbir Singh Pannu. "Efficient Landsat image fusion using fuzzy and stationary discrete wavelet transform." *The Imaging Science Journal* 65, no. 2 (2017): 108-114.
- [14] S.W. Athukorala, L.S. Weerasinghe, M. Jayasooria, D. Rajapakshe, L. Fernando, M. Raffeeze, N.P. Miguntanna, "An analysis of Activity recognition of elder people Variation In Kelani River, Sri Lanka Using Principal Component Analysis," *SAITM Research Symposium on Engineering Advancements* 2013.
- [15] T. Wijesinghe, "Status of Activity recognition of elder people of Kelani River, Central Environmental Authority, Sri Lanka," *IIERR online publication* pp.255, 2010.
- [15] Y. Park, K. H. Cho, J. Park, S. M. Cha, and J. H. Kim, "Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea.," *Sci. Total Environ.*, vol. 502, pp. 31–41, Jan. 2015.
- [16] S. Maiti and R. K. Tiwari, "A comparative study of artificial neural networks, Bayesian neural networks and adaptive neuro-fuzzy inference system in groundwater level recognition," *Environ. Earth Sci.*, vol. 71, no.7, pp. 3147–3160, 2013.
- [17] M.J. Diamantopoulou, V.Z. Antonopoulos and D.M. Papamichail "The Use of a Neural Network Technique for the Recognition of Activity recognition of elder people Parameters of Axios River in Northern Greece",

Journal OfOperational Research, Springer-Verlag, Jan 2005, pp. 115-125.

- [18] Singh, Dilbag, and Vijay Kumar. "Modified gain intervention filter based dehazing technique." *Journal of Modern Optics* 64, no. 20 (2017): 2165-2178.
- [19] Singh, Dilbag, and Vijay Kumar. "Dehazing of remote sensing images using improved restoration model based

dark channel prior." *The Imaging Science Journal* (2017): 1-11.

- [20] Singh, Dilbag, and Vijay Kumar. "Dehazing of Remote Sensing Images using Fourth-order Partial Differential Equations Based Trilateral Filter." *IET Computer Vision* (2017).