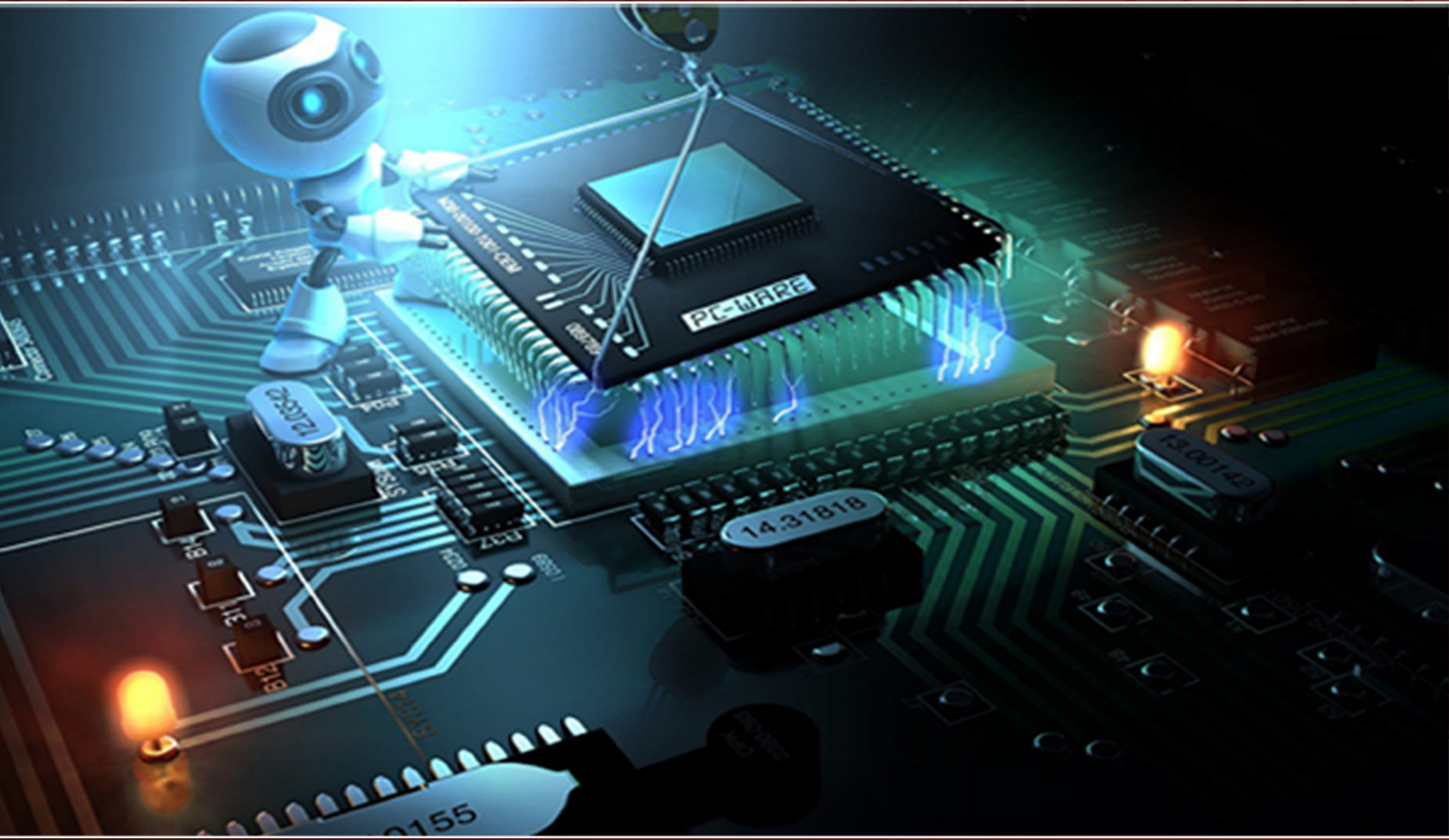
ISSN: 2054-7390

Volume 4 Issue 2, April 2016

Transactions on Machine Learning and Artificial Intelligence



Society for Science and Education
UNITED KINGDOM

TABLE OF CONTENTS

EDITORIAL ADVISORY BOARD

DISCLAIMER

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON
MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
ISSN 2054-7390

A Coordinate Transformation Method based on the Random Variable for Showing the Optical Properties

Jiang Xing-fang[1, 2], Li Xin-lu[1], Kong Xiang-min[1] and Huang Shuang[1, 2]
[1]School of Mathematics and Physics, Changzhou University, China
[2]State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, State Oceanic Administration, Hangzhou, China
xfjiang@cczu.edu.cn

ABSTRACT

A coordinate transformation method of square projection circular with random various has been proposed for the optical properties of the cloaking material. The square produced randomly is projected to the circle and all the points in the perfectly matched layer are transformed. The perfectly matched layer is between the produced square and the border of the square. The first kind transformation has three steps. The first step is that the square grids are produced. The second step is the lattice grids are transformed. The third step is that the lattice grids and those transformed lattice grids are shown. The second kind transformation has four steps. The first step is that the ray came through the center point is assumed. The second step is that the 10 points each ray in the perfectly matched layer are produced randomly. The third step is the 10 points each ray are transformed. The fourth step is that the 10 points each ray and those transformed points are shown. The results show that the transformed results are nothing to do with the position of the randomly produced square. The transformation method is suit for the research of the optical properties in cloaking material.

***Keywords*:** Cloaking material; Coordinate transformation; Projection

1 Introduction

In the past few decades, the metal sandwich were designed with nano-scale to change the distribution of the conductivity and the magnetic permeability. These efforts promote the development of the transformation optics. In particular, Pendry et al[1] and Leonhardt[2] put forward the cloaking material independently since 2006 in Science. The research of the cloaking material has become the hot topic[3-5].

The electromagnetic metamaterial [4] which has cloaking region with micron scale and a high degree of symmetry, such as spherical symmetry, cylindrical symmetry or cubic symmetry. The coordinate transformation [6-7] is important for searching the cloaking regions of the electromagnetic metamaterial. Based on the coordinate transformation invariance for Maxwell's equations the perfectly matched layer [8-9] could be transformed.

2 Coordinate transformation

The model of the researched problem is shown in figure 1[10]. The centers of the square which the side length is 2 times of *a* and of the circle whose radius is *b* are superposition.

DOI: 10.14738/tmlai.42.1811
Publication Date: 06th March, 2016
URL: http://dx.doi.org/10.14738/tmlai.42.1811

The square which the side length is 2 times of *c* is selected randomly and the number of *c* is between *b* and *a*. The center point O of the square which the side length is 2 times of *c* is superposition with the centers of the square which the side length is 2 times of *a* and of the circle which the radius is *b*. The shadow region in figure 1 is between the squares of the side length is 2 times of *c* and of the side length is 2 times of *a*.

In figure 1 the ray is through the center point *O*. On the ray the point *A* is in the square which side length is 2 times of *a*, the point *C* is in the square which side length is 2 times of *c*, the point *B* is in the circle which radius is b, and the point *P* is randomly in the shadow region. The coordinates of *A*, *B*, *C*, *P* are as *A*(*a*, *ay*), *B*(*b*, *by*), *C*(*c*, *cy*), and *P*(*x*,*y*). The transformation is that the point *C*(*c*, *cy*) is transformed to point *B*(*b*, *by*) and the point *P*(*x*,*y*) is transformed to $P'(x', y')$.

The transformation matrix is as equation (1).

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{a-bx}{a-c} & 0 \\ 0 & \frac{ay-by}{ay-cy} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} bx - c\frac{a-bx}{a-c} \\ by - cy\frac{ay-by}{ay-cy} \end{pmatrix} \quad (1)$$

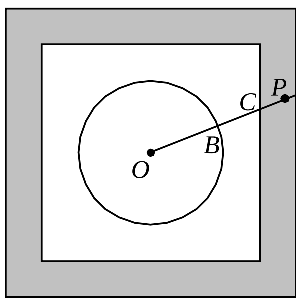


Figure 1 The model of the coordinate transformation

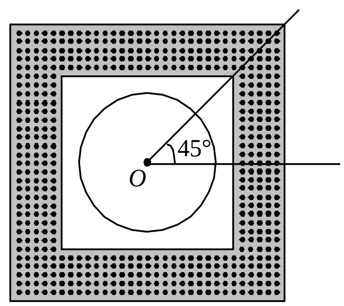


Figure 2 The square grid points

2.1 The transform of the square grid points

In figure 1 the position of the square whose side length is 2 times of c is random and the shadow region is random too. In figure 2 the length of the grid is (*a*-*c*)/6. The larger of the c is, the more the square grid points are. In the shadow region with 45 degrees circular angle the number of the grid point is n1, n2, n3, n4, n5 from left to right. The values of n1, n2, n3, n4, n5 are $\mathrm{round}\left[6*\frac{\mathrm{a}}{\mathrm{a-c}}\right]-5$, $\mathrm{round}\left[6*\frac{\mathrm{a}}{\mathrm{a-c}}\right]-4$, $\mathrm{round}\left[6*\frac{\mathrm{a}}{\mathrm{a-c}}\right]-3$, $\mathrm{round}\left[6*\frac{\mathrm{a}}{\mathrm{a-c}}\right]-2$, $\mathrm{round}\left[6*\frac{\mathrm{a}}{\mathrm{a-c}}\right]-1$ respectively. The flow diagram of the coordinate transformation is shown in figure 3. The coordinate transformation whose circular angle cover 0 to 360 degrees is by the rotation and the inversion based on the transformation results which is cover 0 to 45 degrees of the circular angle .

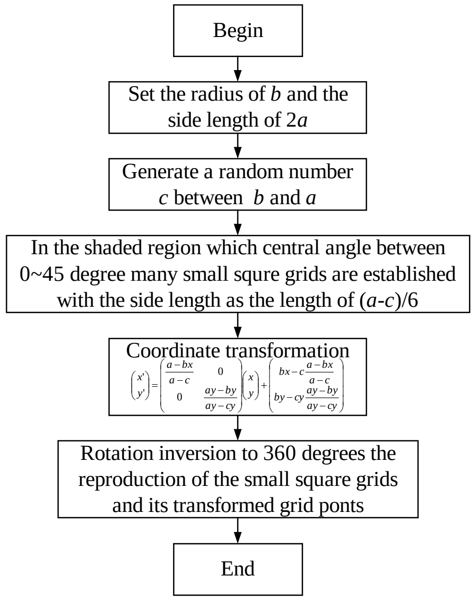
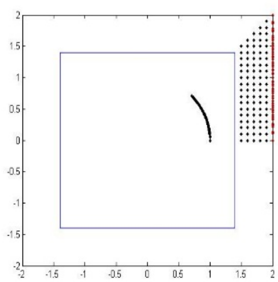
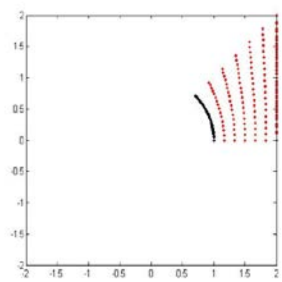


Figure 3 The flow diagram of the coordinate transformation of the square grid points

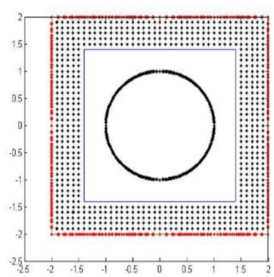
The programming is assumed *a*=2, *b*=1, and *c* is between 1 and 2. The figures of the square grid points and its transformed points are shown in figure 4. The figure 4(a) is the square grid points in 45 degrees for *c*=1.25. The figure 4(b) is the transformed square grid points in 45 degrees for *c*=1.25. The figure 4(c) is the square grid points in 360 degrees for *c*=1.25. The figure 4(d) is the transformed square grid points in 360 degrees for *c*=1.25. The figure 4(e) is the square grid points in 360 degrees for *c*=1.7. The figure 4(f) is the transformed square grid points in 360 degrees for *c*=1.7.



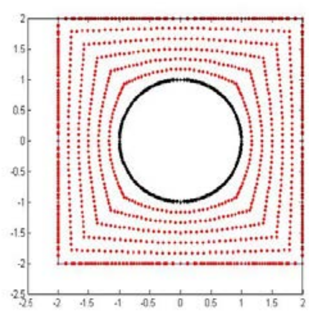
(a) The square grid points in 45 degrees for *c*=1.25



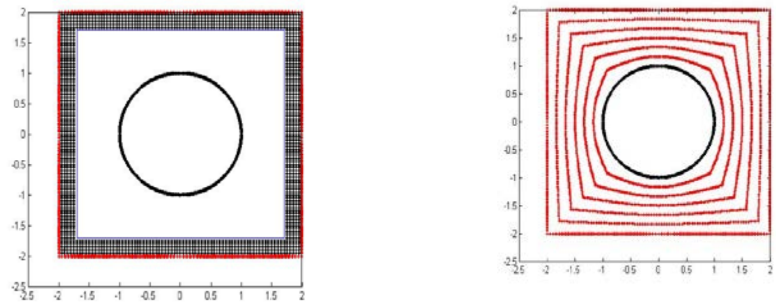
(b) The transformed square grid points in 45 degrees for *c*=1.25



(c) The square grid points in 360 degrees for *c*=1.25



(d) The transformed square grid points in 360 degrees for *c*=1.25



(e) The square grid points in 360 degrees for *c*=1.7 **(f) The transformed square grid points in 360 degrees for *c*=1.7**

Figure 4 The transform of the square grid points

The transformed results shown that the shape made of the transformed points is the same. The difference between the figure 4(d) and the figure 4(f) is the points are different on same shape curve. The number of the points in figure 4(f) is more than that of the points in figure 4(d) because the side length of the grid is small in figure 4(f).

2.2 The transformation of the random points on the ray

In figure 1 the ray is through the center point *O*. The 10 points are chosen randomly on the ray in the shadow region. The transformation is that the point *C* is transformed to point *B*. The point C is in the border of the produced square whose side length is 2 times of *c*. The point *B* is on the circle whose radium is *b*. The points *C* and *B* are on the ray. The point *P* has 10 positions corresponding to 10 random points is transformed to point *P*′. The transform is according to the equation (1) and it can be expressed as the equation (2) too.

$$\overline{OP'} = \frac{\overline{OA}}{\overline{OB}}\overline{CP} \tag{2}$$

The programming is assumed *a*=2, *b*=1, and *c* is between 1 and 2. The flow diagram is shown in figure 5.

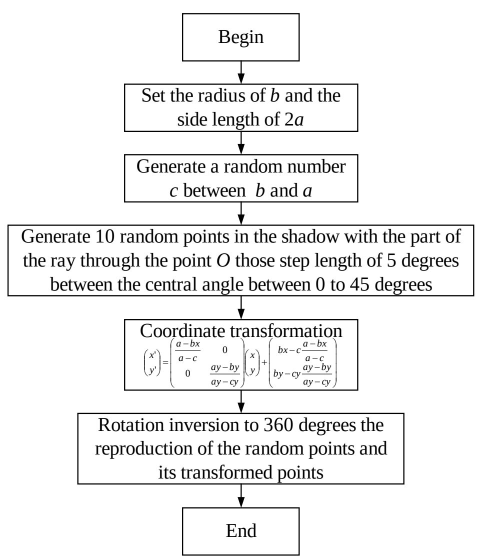
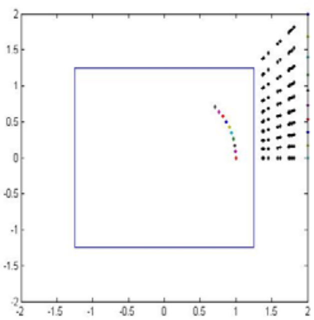
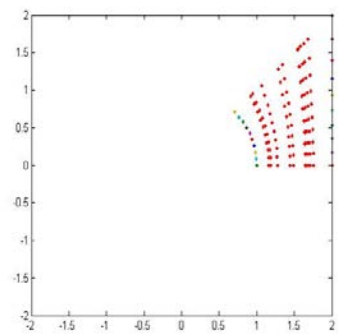


Figure 5 The flow diagram of the coordinate transformation of the random points on a ray

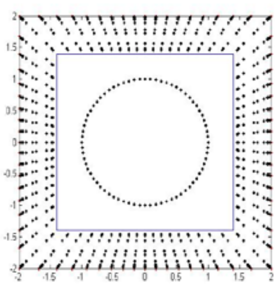
After the square which side length is 2 times of *c* is produced randomly there are four steps in the transform of the random points on a ray. The value of *c* is between *a* and *b*. The first step is to produce the ray which is through the center point *O*. The second step is that the 10 points each ray in the perfectly matched layer are produced randomly. Here the perfectly matched layer is the region between the squares of the side length is 2 times of *c* and of the side length is 2 times of *a*. The third step is the 10 points each ray are transformed according to the equation (1). The fourth step is that the 10 points each ray and those transformed points are shown as figure 6.



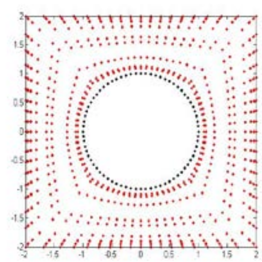
(a) 10 random points each ray in shadow region in 45 degree for c=1.5



(b) 10 transformed points each ray in shadow region in 45 degree for c=1.5



(c) 10 random points each ray in shadow region in 360 degree for c=1.78



(d) 10 transformed points each ray in shadow region in 360 degree for c=1.78

Figure 6 The transform of the random points on the ray

3 Conclusion

The perfectly matched layer is produced after the square whose side length is 2 times of randomly produced *c*. The value of *c* is between *a* and *b*. The outer square which side length is 2 times of *a* and the inner circle whose radium is *b*. The transform is that all the points in the square which side length is 2 times of *c* are transformed to the corresponding points in circle which radium is *b* and the points in the perfectly matched layer are transformed to corresponding points. The results are shown that the transformed square grid points are on the curves of the determined shape and the curves have nothing to do with the position of the square which side length is 2 times of *c* for the square grid points in perfectly matched layer. The random points those in the perfectly matched layer and on the ray which is through the center point *O* are transformed with simple programming. The coordinate transformation method is advantageous to the research of the optical properties in cloaking effect of electromagnetic metamaterial.

ACKNOWLEDGMENT

The authors wish to thank for open issues of State Key Laboratory of Satellite Ocean Environment Dynamics (No. SOED1503) and of Jiangsu Province Key Laboratory of Modern Optical Technologies (No. KJS1405).

REFERENCES

[1] J. B. Pendry, D. Schurig, D. R. Smith. Controlling electromagnetic fields [J]. Science. 2006, 312:1780-1782

[2] Ulf Leonhardt. Optical conformal mapping and dielectric invisibility devices [J]. Science. 2006, 312: 1777-1780

[3] Adrian Cho. High-tech materials could render objects invisible [J]. Science. 2006, 312: 1120

[4] Raimond Grimberg. Electromagnetic metamaterials [J]. Materials Science and Engineering B. 2013, 178: 1285-1295

[5] Jaeyub Hyun, Junghwan Kook, Semyung Wang. Efficient and stable model reduction scheme for the numerical simulation of broadband axoustic metamaterials [J]. Computers and Mathematics with Application. 2015, 69: 876-892

[6] D. Schurig, J. J. Mock, B. J. Justice, et al. Metamaterial electromagnetic cloak at microwave frequencies [J]. Science. 2006, 314: 977-980

[7] Ozlem Ozgun, Mustafa Kuzuoglu. Monte Carlo simulations of Helmholtz scattering from randomly positioned array of scatters by utilizing coordinate transformations in finite element method [J]. Wave Motion.2015, 56: 165-182

[8] Archana Rajput, Kumar vaibhav Srivastava. Dispersion study on scattering cross section of metamaterial cloak due to various cloaking parameters [J]. Optik. 2015, 126: 2362-2367

[9] ZiyuWang, WeiWei, NiHu, et al. Manipulation of elastic waves by zero index metamaterials [J]. Journal of Application Physics. 2014, 116; 204501,1-5

[10] Ozlem Ozgun, Mustafa Kuzuoglu. Electromagnetic metamorphosis: reshaping scatters via conformal anisotropic metamaterial coatings [J]. Microwave and Optical Technology Letters. 2007, 49: 2386-2392

SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE ISSN 2054-7390

Fuzzy Rough Classification Models for Network Intrusion Detection

Ashalata Panigrahi[1] and Manas Ranjan Patra[2]
Department of Computer Science, Berhampur University, Berhampur, India
[1]ashalata.panigrahi@yahoo.com; [2]mrpatra12@gmail.com

ABSTRACT

In recent years advancements in communication technology have led to a wide range of Internet services. While an overwhelming number of Internet users have shown interest in such services, incidences of cyber-attacks by miscreants have thwarted their dependence on electronically-accessible services. In order to deal with this alarming situation intrusion detection systems (IDS) have emerged as a potential solution to analyse network activities of users and report attempts of possible intrusions**.** Building an intrusion detection system is a complex and challenging task. This requires analysis of network data from several dimensions so as to develop a pragmatic system to handle different forms of intrusive behaviour of attackers. In this paper, we propose a hybrid intrusion detection approach which combines techniques based on both fuzzy and rough set theories to classify network data as normal and anomalous. Our approach comprises of two phases; in the first phase the most relevant features are extracted using a set of rank and search based methods; and in the second phase we classify the reduced dataset as normal or anomalous using five different classifiers, namely, Fuzzy Nearest Neighbour, Fuzzy-Rough Nearest Neighbour, Fuzzy-Rough Ownership NN, Vaguely Quantified Nearest Neighbours, and Ordered Weighted Average Nearest Neighbours. Experimental results show that the proposed hybrid approach has the ability to achieve high intrusion detection rate and low false alarm

Keywords: FNN, Fuzzy-Rough NN, FRONN, VQNN, OWANN.

1 Introduction

The last decade has witnessed an unprecedented expansion in Internet connectivity which has led to a plethora of internet based services catering to a wide range of user groups. This has evoked security concerns for protecting personal and sensitive data from misuse. As more and more number of users get connected to internet, the window of opportunity for malicious users to fiddle with user data becomes lucrative. Network security deals with the confidentiality, integrity, availability and protection of data as well as computing resources. Different approaches have been adopted to implement a range of security measures such as authentication, cryptography, firewalls, antivirus, spywares, Virtual Private Network, and intrusion detection systems (IDS) but none of them is capable of providing complete security. Malicious users constantly look for ways to by-pass the security features, and many-a-times succeed in accessing important network resources. As a result developing flexible and adaptive security systems is a major challenge. In this context, IDSs are becoming important tools to ensure network security where IDSs

DOI: 10.14738/tmlai.42.1882
Publication Date: 15th Mar, 2016
URL: http://dx.doi.org/10.14738/tmlai.42.1882

are deployed to dynamically monitor all incoming and outgoing network activities taking place in a system and distinguish between legitimate and anomalous network users. Hybrid IDS are dynamic defensive systems, capable of adapting to dynamically changing traffic patterns and try to detect varieties of network attacks.

2 Related Work

Classification techniques are being used to build predictive models in different application domains. Network intrusion detection is one such area which extensively uses different classifiers to build predictive models to distinguish between intrusions and normal connection requests in a network setup. Several works have been reported utilizing different classification techniques to analyse intrusion data and build prediction models with the sole objective of enhancing intrusion detection accuracy and lowering false alarms.

Gong, S [5] has proposed a feature selection approach based on Genetic Quantum Particle Swarm Optimization (GQPSO) for network intrusion detection wherein selection and variation of genetic algorithm with QPSO algorithm have been combined to reduce redundant and irrelevant features. Experimental results show that the GQPSO algorithm performs better than PSO and QPSO algorithms in terms of detection rate and speed of classification. Hoque et al**.** [6] have implemented an Intrusion Detection System by applying genetic algorithms to efficiently detect various types of network intrusive activities. To measure the efficiency of their system they used the standard KDD 99 intrusion detection benchmark dataset and obtained realistic detection rate. But their performance of detection rate was poor and they failed to reduce the false positive rate. Zhou et al. [7] presented a hierarchical neuro-fuzzy inference intrusion detection system (HFIS). In their proposed system, principal component analysis neural network was used to reduce the input data space. An enhanced fuzzy c-means clustering algorithm was applied to create and extract fuzzy rules. The adaptive neural fuzzy inference system was utilized repeatedly in their model. At last, the system was optimized by genetic algorithm. The main advantages of the HFIS model are its capability to perform not only misuse detection but also anomaly detection. Moreover, their method has higher speed and better performance.

Tong et al. [8] have proposed a hybrid IDS based on RBF/Elman neural network wherein the RBF neural network is employed as a real time pattern classifier while Elman neural network is employed to restore the memory of past events. Mohamadi H [9] proposed Simulated Annealing (SA) based fuzzy system to develop an Intrusion Detection System (IDS). The use of SA in IDS is an attempt to effectively explore and exploit the large search space associated with intrusion detection classification problem. Experiments were carried out on 10% of KDD Cup99 dataset of UCI KDD archive. Due to the imbalanced records in the dataset a subset of the dataset was used as training and testing sets (20752 randomly generated samples) and normalized between 0.0 and 1.0. Initial set of fuzzy if-then rules was generated and initial temperature was set as 100. The fitness of the rule was evaluated by number of correctly classified training patterns. The results showed that average accuracy rate obtained was varying from 94% to 99% with the number of rules ranging from 50 to 100. This approach was compared with the different baseline classifiers including pruning C4.5, Naïve Bayes, K-NN, SVM and multi-objective genetic fuzzy IDS. The results showed that the proposed approach obtained highest accuracy (92.89%), better precision, lowest classification cost (0.2093), F-measure, recall than other classifiers. In our previous work [10] we have proposed a hybrid classification model based on evolutionary computation based techniques. The result shows that AIRS1 classifier with best first search feature selection gives highest accuracy and AIRS2 classifier with Gain Ratio feature selection gives lowest false alarm rate.

3 Proposed Hybrid Intrusion Detection Model

The aim of this work is to build a high performance hybrid intrusion detection model that can achieve low false alarm rate and high detection rate. The model comprises of two levels as depicted in figure 1.

Level-1consists of feature selection methods to extract the most relevant features from the intrusion dataset which can contribute to the classification process. This is achieved by identifying the irrelevant and redundant information in the intrusion dataset and discarding them from the dataset. Four different rank methods namely, Gain Ratio, Relief-F, One-R, Symmetrical Uncertainty and three different search methods namely, Best First, Greedy Stepwise, Rank Search have been applied for selection of relevant attributes. At Level-2 the reduced data obtained from Level-1 is classified using five classification techniques namely, Fuzzy Nearest Neighbour, Fuzzy-Rough Nearest Neighbour, Fuzzy-Rough Ownership NN, Vaguely Quantified Nearest Neighbours, and Ordered Weighted Average Nearest Neighbours. The NSL-KDD dataset has been used for building and validating the model. Further, 10-Fold cross-validation has been employed for analysis of detection rate, accuracy, false alarm rate, and fitness value.

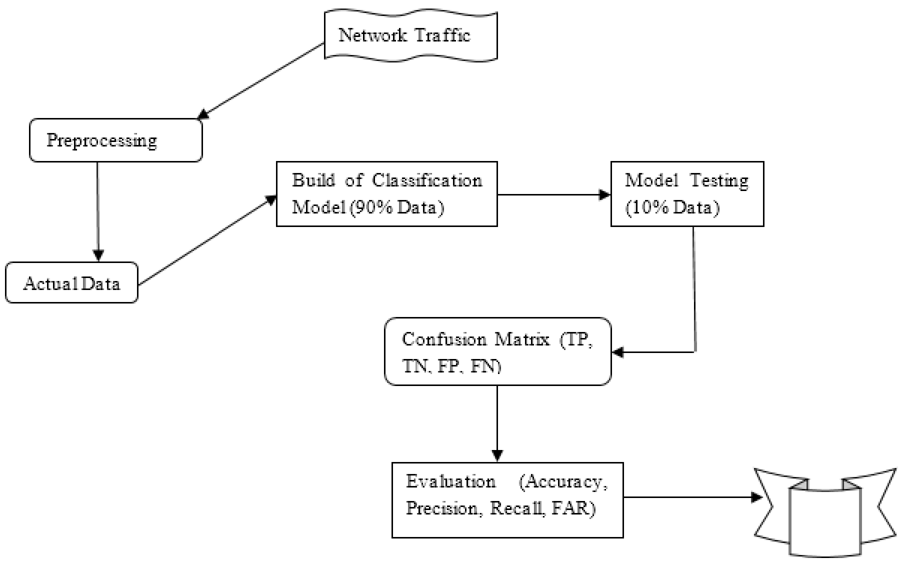


Figure. 1 System diagram for Hybrid IDS

4 Methodology

4.1 Hybridization of Rough Set and Fuzzy Set

Fuzzy Set

A fuzzy set [4] in X is an $X \rightarrow [0, 1]$ mapping, while a fuzzy relation in X is a fuzzy set in $X \times X$. For all y in X, the R-forest of y is the fuzzy set R_y is defined by

$$R_y(x) = R(x,y) \tag{1}$$

For all x in X, if R is reflexive and symmetric fuzzy relation, that is

$$R(x,x) = 1 \tag{2}$$

$$R(x,y) = R(y,x) \quad (3)$$

holds for all x and y in X, then R is called a “fuzzy tolerance ratio”.

Rough Set

Rough Set Theory is a mathematical tool to deal with imprecise and insufficient knowledge [3]. In rough set theory, membership is not the primary concept unlike fuzzy sets. It deals with inconsistency, uncertainty, and incompleteness by imposing an upper and a lower approximation to set membership. The advantage of rough set theory is that it does not require any preliminary or additional information about data, like probability in statistics or grade of membership/value of possibility in fuzzy set theory.

Let (X, A) be an information system where X is the universe of discourse and A is a non-empty finite set of attributes such that $a: X \rightarrow V_a$ for every $a \in A$. The set V_a is called the “value set of a”. Given $B \subseteq A$ there is an associated equivalence relation R_B:

$$R_B = \{ (x,y) \in X^2 \mid \forall a \in B, a(x) = a(y) \} \quad (4)$$

If $(x,y) \in R_B$, then x and y are indiscernible by attributes from B. The equivalence classes of the B-indiscernibility relation are denoted by $[x]_B$.

Let A be a subset X. A can be approximated using the information contained within B by constructing the B-lower and B-upper approximations of A.

$$R_B{\downarrow}A = \{ x \in X \mid [x]_B \text{ subset } A \} \quad (5)$$

$$R_B{\uparrow}A = \{ x \in X \mid [x]_B \cap A \neq \emptyset \} \quad (6)$$

The tuple $(R_B{\downarrow}A, R_B{\uparrow}A)$ is called a rough set.

Fuzzy-Rough Set Theory

Hybridizing fuzzy rough set theory is focused mainly on fuzzifying the formulas for lower and upper approximations [2]. Given a fuzzy tolerance relation R and a fuzzy set A in X, the lower and upper approximation of A by R can be defined as:

$$(R{\downarrow}A)(x) = \inf_{y \in X} I(R(x,y),A(y)) \quad (7)$$

$$(R{\uparrow}A)(x) = \sup_{y \in X} T(R(x,y),A(y)) \quad (8)$$

Here *I* is an implicator and *T* is a t-norm.

4.2 Fuzzy Nearest Neighbour Classification

The Fuzzy Nearest Neighbour (FNN) algorithm [11] was introduced to classify test objects based on their similarity to a given number K of neighbours, and these neighbours’ membership degree to (crisp or fuzzy) class labels. For the purpose of (FNN), the extent *C’(y)* to which an unclassified object y belongs to a class *C* is computed as:

$$C'(y) = \sum_{x \in N} R(x,y)\, C(x) \quad (9)$$

where N is the set of object y's K nearest neighbours, and *R(x,y)* is the [0,1]-valued similarity of *x* and *y*.

The Fuzzy K-Nearest Neighbour Algorithm

```
   FNN (X, C_D, y, K)
Input:  X: the training data set; C_D:  the set of decision classes;
        y:  the objects to be classified; K: the number of nearest neighbours
1. begin
2.     N ← get Nearest Neighbours ( y, K )
3      for each C ∈ C_D  do
4.         C' (y) = Σ_{x∈N} R(x, y) C(x)
5.     end
6. end
   Output: arg max ( C' (y) )
```

4.3 Fuzzy-Rough Nearest Neighbour Classification

In Fuzzy-Rough Nearest Neighbour (FRNN) algorithm the nearest neighbours are used to construct the fuzzy lower and upper approximations of decision classes, and test instances are classified based on their membership to these approximations. FRNN algorithm combines fuzzy-rough approximation with the classical FNN approach [12]. The rationale behind the algorithm is that the lower and upper approximation of a decision class, calculated by means of the nearest neighbours of a test object y, provides good clues to predict the membership of the test object to that class. The algorithm is dependent on the choice of a fuzzy tolerance relation R. Given the set of conditional attributes A, the fuzzy tolerance relation R is defined by

$$R(x,y) = \min_{a \in A} R_a(x,y) \tag{10}$$

in which $R_a(x,y)$ is the degree to which objects *x* and *y* are similar for attribute a. Here we choose

$$R_a(x,y) = 1 - \frac{|a(x) - a(y)|}{|a_{max} - a_{min}|} \tag{11}$$

If (R↓C)(y) is high , it reflects that all of y's neighbours belong to C. A high value of (R↑C) means that at least one neighbour belongs to that class.

The Fuzzy Rough Nearest Neighbour Algorithm:

```
FRNN ( X, C_D, y)
Input: X: the training data set;  C_D, the set of decision classes;
       y: the objects to be classified;
1.begin
2.     N ← get Nearest Neighbours ( y, K )
3.     τ←  0 , Class ← Ø
4.     for each  C ∈  C_D  do
5.         . if(( R↓C )(y) +  ( R↑C)(y)) / 2  ≥  τ then
6.                 Class ← C
7.          τ  ←  (( R↓C )(y) +  ( R↑C)(y)) / 2
8.         endif
9 .    end
10. end

Output Class
```

4.4 Fuzzy-Rough Ownership NN Classification

A fuzzy-Rough ownership is an attempt to handle both "fuzzy uncertainty" caused by overlapping classes and "rough uncertainty" caused by insufficient knowledge [12]. All training objects influence the ownership function. The algorithm does not use fuzzy lower or upper approximations to determine class membership. The fuzzy-rough ownership function τ_c of class *C* for an object *y* is defined as,

$$\tau_c(y) = \sum_{x \in X} \frac{R(x,y)C(x)}{|X|} \tag{12}$$

where the fuzzy relation R is determined by

$$R(x,y) = \exp(-\sum_{a \in A} Ka(a(y) - a(x))2/(m-1)) \tag{13}$$

where m controls the weighting of the similarity and K_a is a parameter that decides the bandwidth of the membership and K_a is defined as

$$K_a = \frac{|X|}{2 \sum_{x \in X} \| a(y) - a(x) \| 2/(m-1)} \tag{14}$$

$\tau_c(y)$ is interpreted as the confidence with which y can be classified to class C.

```
FROWNN (X, A, C_D, y )
Input: X the training data set; A the set of conditional features;
       C_D the set of decision classes; y the object to be classified.
begin
     for each a ∈ A do
          K_a = |X| / (2 Σ_{x∈X} || a(y) − a(x)||2/(m−1))
     end
     N ← | x |
     for each C ∈ C_D do τ_c(y) = 0
          for each x ∈ N do
               d = Σ_{a∈A} Ka (a(y) − a(x))^2
          for each C ∈ C_D do
               τ_c(y)+ = C(x). exp (− d^{1/(m−1)} ) / | N |
          end
     end
  end
end
Output   argmax_{C ∈ C_D} τ_c(y)
```

4.5 Vaguely Quantified Nearest Neighbours Classification

VQNN [12] depends only on the summation of the similarities of each class. It uses the linguistic quantifiers "most" and "some". Given a couple (Q_u, Q_l) of fuzzy quantifiers that represent "most" and "some" respectively, the lower and upper approximation of C. VQNN assigns a class to a target instance y as follows:

i. Determine NN, the K nearest neighbours of *y*.
ii. Assign y to the class *C* for which $(R\downarrow^{Q_u} C)(y) + (R\uparrow^{Q_l} C)(y)$ is maximal.

The upper and lower approximation of Vaguely Quantified rough sets are defined as

$$((R\downarrow^{Qu} C)(y)) = Q_u\left(\frac{\sum_{x \in X} \min(Rx,y),C(x)}{\sum_{x \in X} R(x,y)}\right) \tag{15}$$

$$((R\uparrow^{Ql} C)(y)) = Q_l\left(\frac{\sum_{x \in X} \min(Rx,y), C(x)}{\sum_{x \in X} R(x,y)}\right) \quad (16)$$

The operators Q_u and Q_l are fuzzy quantifiers that represent most and some respectively. They are increasing [0, 1] → [0, 1] mapping such that

$$Q_u(1) = Q_l(1) = 1 \text{ and } Q_u(0) = Q_l(0) = 0$$

This classifier based on rough set theory is capable of handling noise data.

4.6 Ordered Weighted Average Nearest Neighbours Classification

The OWA operator [13] models an aggregation process in which a sequence A of n scalar values are ordered decreasingly and then weighted according to their ordered position by a weighting vector W = {w1, w2, ……,wp}. The OWA_W operator aggregates p values A = { a1, a2, ………., ap } as follows:

$$OWA_W(a1, a2, \ldots\ldots, ap) = \sum_{i=1}^{p} wi\ bi \quad (17)$$

where bi= aj if aj is the i-th largest value in A = {a1, a2, ………., ap}.

The weights W are associated with ordered positions. The higher values in {a1, a2, ……….,ap} are assigned to the first weights in W and the lower values are associated with the last weights in W.

Let R be a fuzzy relation in X and A a fuzzy set in X = { x1, x2, ……, xn}. Let Ʈ be a t-norm and I, a fuzzy implication. The OWA-based lower and upper approximation of A under R with weight vectors W_l and W_u are defined as

$$(R\downarrow W_l\ A)(y) = OW\ A_{Wl}(I(R(x_i, y), A(x_i))) \quad (18)$$

$$(R\uparrow W_u\ A)(y) = OW\ A_{Wu}(I(R(x_i, y), A(x_i))) \quad (19)$$

5 Experimental Setup

5.1 NSL-KDD Dataset

NSL- KDD is a dataset proposed by Tavallace et al. [14] which is a reduced version of the original KDD'99 dataset. NSL-KDD consists of same features as KDD'99 training dataset but has the following advantages over the original KDD'99dataset.

a) The training set does not include redundant records.
b) The test set has no duplicate records.
c) The number of records in the training and test set is reasonable, which makes it affordable to run experiments on the complete set without the need to randomly select a small portion. Consequently, the evaluation of results reported by different researchers can be comparable.

The data set consists of 41 feature attributes out of which 38 are numeric and 3 are symbolic. Total number of records in the data set is 125973 out of which 67343 are normal and 58630 are attacks. The dataset contains different attack types that could be classified into four main categories namely, Denial of Service (DOS), Remote to Local (R2L), User to Root (U2R), and Probing

The percentage distribution of records under each category of attack is provided in Table 1 and figure 2.

Table 1 Data Distribution of NSL-KDD Dataset

Class	Number of Records	% of occurrence
Normal	67343	53.48%
DOS	45927	36.45%
R2L	995	0.78%
Probes	11656	9.25%
U2R	52	0.04%
Total	125973	100%

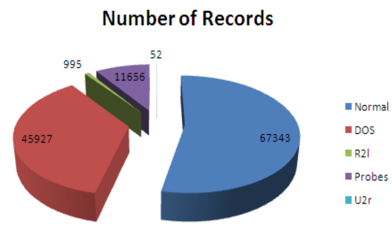


Figure. 2 Distribution of Records

5.2 Feature Selection

In order to build a high performance IDS, selection of the most relevant features present in the intrusion dataset is an important research challenge. Feature selection can be defined as a process that chooses a minimum subset of M features from the original set of N features, so that the feature space is optimally reduced according to certain evaluation criteria. As the dimensionality of a domain expands, the number of features N increases. Finding the best feature subset is usually intractable [15].

Feature selection improves classification performance by searching for the subset of features, which best classifies the training data. In case of high dimensional feature space, some of the features may be redundant or irrelevant. Removing these redundant or irrelevant features is very important as they may deteriorate the performance of classifiers. Feature selection involves finding a subset of features from the dataset, thereby decreasing the size of the original dataset in order to improve prediction accuracy of the classifier [16]. Now, we present the feature selection techniques that we have applied for reducing the NSL-KDD dataset with the most desirable features which can improve the performance of the classifiers.

Gain Ratio

The information gain measure prefers to select features having a large number of values. The extension of information gain is known as gain ratio [17] and is based on ranking which attempts to overcome any bias. It applies a kind of normalization to information gain using a "split information" value. The split information value represents the potential information generated by splitting the training dataset D into v partitions, corresponding to v outcomes on attribute A

$$\text{SplitInfo}_A(D) = -\sum_{j=1}^{v} |D_j| / |D| * \log_2(|D_j| / |D|) \quad (20)$$

This value represents the potential information generated by splitting the training dataset D into v partitions corresponding to the v outcomes of a test on attribute A.

The gain ratio is defined as

$$\text{GainRatio (A)} = \text{Gain (A)} / \text{SplitInfo(A)} \quad (21)$$

The feature with the maximum gain ratio is selected as the splitting attribute.

One-R

One-R (short for One Rule) algorithm proposed by Holte [18] is a simple classification algorithm that generates a one-level decision tree expressed in the form of a set of rules all of which test one particular feature. It is capable of generating good rules for characterizing the structure in data.. One-R can handle missing values and numeric features. The One-R algorithm generates rules and tests a single feature at a time and a branch for every value of that feature. For every branch, the class with the best classification is selected.

Relief-F

Relief-F feature selection method is one of the most successful algorithms for assessing the quality of features due to its simplicity and effectiveness. Relief-F can handle noise and multiclass datasets [19]. Relief-F feature evaluation [20] evaluates the worth of a feature by repeatedly sampling an instance and considering the value of the given feature for the nearest instance of the similar and different classes. This feature evaluation assigns a weight to each feature based on the ability of the feature to distinguish among the classes, and then selects those features whose weights exceed a user-defined threshold as relevant features. The three basic steps of Relief-F feature evaluator technique are:

- Calculate the nearest miss and nearest hit
- Calculate the weight of a feature
- Return a ranked list of features or the top K features according to a given threshold

The function diff (Feature, Instance1, Instance2) computes the difference between the values of a feature for two different instances. For discrete attributes the difference is either 1 (the values are different) or 0 (the values are the same), whereas for continuous features the difference is the actual difference normalized to the interval [0, 1]. Higher the value of m (the number of instance sampled), the more reliable is Relief-F’s estimate.

Symmetrical Uncertainty

Symmetrical uncertainty technique [17] is symmetric in nature and it reduces the number of comparisons required. It is not influenced by multi-valued features and its values are normalized to the range [0, 1]. This technique consists of two phases to select the most informative features to target classes from the original feature space. In the first phase (lines 1-5 in the algorithm), irrelevant features with poor prediction ability to target a class are removed. In the second phase (lines 7-12 in the algorithm) redundant features that are inter-correlated with one or more of other features are eliminated.

Given a dataset with a number of input features and a target class, the algorithm first calculates the mutual information between features and class. The algorithm then ranks the features in descending order according to their degrees of association to the target class. Once the input features are ranked, those terms whose information measures are greater than zero are kept; which means the removed features are totally irrelevant to target class and the remaining ones are predictive. Next, it starts by calculating the inter-correlated strengths of each pair of features. The total amount of mutual information for each feature is acquired by adding all mutual information measures together that relate to the feature.

Best First Search

Best First Search (BFS) [21] uses classifier evaluation model to estimate the merits of features. The feature with high merit values are considered as potential features and thus selected for classification. Best first moves through the search space by making local changes to the current feature subset. It searches the space of feature subsets by augmenting with a backtracking facility. Given enough time, a best first search will explore the entire search space, thus it is common to use a stopping criterion. It may start with an empty set of features and search forward, or start with the full set of features and search backward, or start at any point and search in both the directions.

Greedy Stepwise Search

Greedy Stepwise search [21] performs a greedy forward or backward search through the space of feature subsets. It may start with no / all features or from an arbitrary point in the space and stops when addition/ deletion of any feature results in decrease in evaluation. This can also produce a ranked list of features by traversing the space from one side to other and recording the order in which features are selected.

Rank Search

This uses a subset evaluator to rank all features. If a subset evaluator is specified, then a forward selection search is used to generate a ranked list. Next, from the ranked list of features a subset of best feature set is selected. Table 4 enlists the features selected after application of each of the above feature selection technique.

Table 4 Selected Attributes after Feature Selection

Feature Selection Method	No. of Features Selected	Feature Names
Gain Ratio	10	Flag, Src_bytes, Dst_bytes,Logged_in, Serror_rate, Srv_serror_rate, Same_srv_rate, Diff_srv_rate, Dst_host_serror_rate, Dst_host__srv_serror_rate.
One-R	14	Service,Flag, Src_bytes, Dst_bytes, Count,Serror_rate, Srv_serror_rate, Same_srv_rate, Diff_srv_rate, Dst_host_srv_count, Dst_host_same_srv_rate, Dst_host_diff_srv_rate, Dst_host_serror_rate, Dst_host__srv_serror_rate.
Relief Attribute Evaluator	12	Protocol_type, Service, Flag, Count,Same_srv_rate, Dst_host_count, Dst_host_srv_count, Dst_host_same_srv_rate, Dst_host_diff_srv_rate, Dst_host_same_srv_port_rate, Dst_host_serror_rate, Dst_host_rerror_rate
Symmetrical Uncertain Attribute Evaluator	16	Service, Flag, Src_bytes, Dst_bytes, Logged_in, Count, Serror_rate, Srv_serror_rate, Same_srv_rate, Diff_srv_rate, Dst_host_srv_count, Dst_host_same_srv_rate, Dst_host_diff_srv_rate, Dst_host_srv_diff_host_rate, Dst_host_serror_rate, Dst_host__srv_serror_rate.
Best First Search	13	Duration, Service, Src_bytes, Dst_bytes, Logged_in, Count,Ser_rate,Dst_h_co, Ds_ho_sr, Ds_Rate, Ds_d_h_rt, Ds_h_r, Ds_hrr.
Rank Search	13	Service, Flag, Src_bytes, Dst_bytes, Logged_in, Root_shell, Serror_rate, Srv_serror_rate, Same_srv_rate, Diff_srv_rate, Dst_host_srv_diff_host_rate, Dst_host_serror_rate, Dst_host__srv_serror_rate
Greedy Stepwise	11	Service,Flag, Src_bytes, Dst_bytes, Logged_in, Root_shell, Srv_serror_rate,Same_srv_rate, Diff_srv_rate, Dst_host_srv_diff_host_rate, Dst_host_serror_rate,

Cross Validation

Cross validation calculates the accuracy of the model by separating the data into two different populations, a training set and a testing set. In k-fold cross-validation [17] the dataset is randomly partitioned into n mutually exclusive folds, T_1, T_2, …..,T_n each of approximately equal size. Training and testing are performed n times. Each training set consists of (n – 1)/n th of the dataset and the remaining 1/n th is used as test data In 10-fold cross validation, a given dataset is partitioned into 10 subsets. Out of these 10 subsets, 9 subsets are used to perform a training fold and a single subset is retained as testing data. This cross-validation process is then repeated 10 times (the number of folds). The 10 sets of results are then aggregated by averaging to produce a single model estimation. The advantage of 10-fold cross validation over random sub-sampling is that all objects are used for both training and testing, and each object is used for testing only once per fold.

Confusion Matrix

The confusion matrix is a table with two rows and two columns that reports the number of False Positive, False Negative, True Positive, True Negative. The confusion matrix maintains the information about actual and predicted classes. An IDS is evaluated by its ability to make accurate prediction of attacks. Intrusion detection systems mainly discriminate between two classes, attack class (abnormal data), and normal class (normal data). While classifying the attacks and normal access behaviour of users, there can be four possibilities as depicted in Table 5 such as True Positives, False Positives, True Negatives, and False Negatives.

Table.5 IDS Confusion matrix

		Predicted Class	
		Negative Class (Normal)	Positive Class (Attack)
Actual Class	Negative Class(Normal)	True Negative (TN)	False Positive (FP)
	Positive Class (Attack)	False Negative (FN)	True Positive (TP)

The accuracy, detection rate, precision, F-value, false alarm rate, fitness value are calculated as follows

Accuracy measure the probability that the algorithm can correctly predict positive and negative examples and is given by:

$$\text{Accuracy} = \frac{TP+TN}{TN+TP+FN+FP}$$

$$\text{Detection Rate or Recall} = \frac{TP}{TP+FN}$$

Precision is a measure of the accuracy provided that a specific class has been predicted and it is calculated as:

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

F- Value is the harmonic mean of Precision and Recall which measures the quality of classification which is given by:

$$\text{F - Value} = 2 * \frac{(\text{ Precision} * \text{Recall})}{(\text{Precision} + \text{Recall })}$$

False Alarm Rate is defined as the number of normal instances incorrectly labelled as intrusion divided by the total number of normal instances and is given by:

$$\text{False Alarm Rate} = \frac{\mathrm{FP}}{\mathrm{TN+FP}}$$

$$\text{Fitness Value} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} * \frac{\mathrm{TN}}{\mathrm{TN+FP}}$$

6 Results and Discussion

Here, we study the effectiveness of the hybrid intrusion detection model that uses five classification techniques, viz., Fuzzy Nearest Neighbour, Fuzzy-Rough Nearest Neighbour, Fuzzy-Rough Ownership NN, Vaguely Quantified Nearest Neighbours, Ordered Weighted Average Nearest Neighbour along with different feature selection methods. The performance of different combinations of classifiers and feature selection methods are evaluated on the basis of accuracy, detection rate, precision, F-value, false alarm rate, fitness value, and error rate. The results are summarized in Table 6 and Table 7.

Table 6 Comparison of Accuracy, Detection rate, precision, F-value, false alarm rate, fitness value, of five classification techniques using Ranking Attribute Reduction methods

Attribute Reduction Method	Test Mode	Classifier Techniques	Accuracy in %	Detection Rate in %	Precision in %	F-Value in %	False Alarm Rate in %	Fitness Value in %
One-R	10-Fold Cross-Validation	Fuzzy NN	99.2427	99.0159	99.3548	99.185	0.5598	98.4614
		Fuzzy Rough NN	98.9712	98.9749	98.8165	98.8956	1.0320	97.9534
		Fuzzy Ownership NN	**99.4292**	**99.5037**	**99.3139**	**99.4087**	**0.5986**	**98.908**
		VQNN	98.8998	98.7105	98.9231	98.8167	0.9355	97.7871
		OWANN	98.9109	98.7395	98.9184	98.8288	0.934	97.8113
Relief-F	10-Fold Cross-Validation	Fuzzy NN	89.4414	88.7054	88.6193	88.6623	9.9179	79.9077
		Fuzzy Rough NN	99.4753	99.2734	99.5979	99.4354	0.3489	98.9269
		Fuzzy Ownership NN	99.2856	99.34	99.1269	99.2334	0.7618	98.5832
		VQNN	**99.4792**	**99.3809**	**99.4996**	**99.4402**	**0.4351**	**98.9485**
		OWANN	99.4507	99.3553	99.4638	99.409	0.3267	98.892
SU	10-Fold Cross-Validation	Fuzzy NN	99.2935	99.0141	99.3888	99.2011	0.833	98.4892
		Fuzzy Rough NN	99.3499	99.3399	99.2637	99.3018	0.6415	98.7026
		Fuzzy Ownership NN	**99.542**	**99.5173**	**99.5411**	**99.5292**	**0.3996**	**99.1192**
		VQNN	99.2252	99.2922	99.0455	99.1681	0.833	98.465
		OWANN	99.207	99.2819	99.0168	99.1492	0.8583	98.4298
Gain Ratio	10-Fold Cross-Validation	Fuzzy NN	96.8898	98.0641	95.3831	96.705	4.1326	94.0115
		Fuzzy Rough NN	98.941	98.2398	99.4784	98.8528	0.4484	97.7992
		Fuzzy Ownership NN	**99.1609**	**98.6798**	**99.556**	**99.1199**	**0.3832**	**98.3016**
		VQNN	98.9387	98.3575	99.3556	98.8569	0.5554	97.8112
		OWANN	98.9315	98.3592	99.3385	98.8464	0.5702	97.7984

It is observed that Fuzzy ownership nearest neighbour classification technique with symmetrical uncertainty feature selection yields better accuracy and low false alarm rate than other classification techniques. A comparison of classifiers with respect to accuracy, recall / detection rate, and false alarm rate is presented in figures 3, 4, and 5 respectively using rank based feature selection.

Table7 Comparison of Accuracy, Detection rate, precision, F-value, false alarm rate, fitness value of five classification techniques using Searching Attribute Reduction methods

Attribute Reduction Method	Test Mode	Classifier Techniques	Accuracy in %	Detection Rate in %	Precision in %	F-Value in %	False Alarm Rate in %	Fitness Value in %
Best First Search	10-Fold Cross-Validation	Fuzzy NN	99.5594	99.6401	99.4654	99.5526	0.5108	99.1311
		Fuzzy Rough NN	99.5142	99.49	99.4322	99.461	0.4648	99.276
		Fuzzy Ownership NN	**99.5729**	**99.606**	**99.533**	**99.5694**	**0.4071**	**99.2006**
		VQNN	99.3761	99.3485	99.3112	99.3299	0.5999	98.7524
		OWANN	99.3403	99.3263	99.2569	99.2916	0.6473	98.6833
Greedy Stepwise	10-Fold Cross-Validation	Fuzzy NN	95.0473	92.0706	97.1388	94.5352	2.361	89.8967
		Fuzzy Rough NN	99.615	99.4849	99.6872	99.5859	0.2717	99.2145
		Fuzzy Ownership NN	**99.6356**	**99.6145**	**99.6451**	**99.6288**	**0.309**	**99.3067**
		VQNN	99.438	99.3246	99.4671	99.3958	0.4633	98.8644
		OWANN	99.4221	99.316	99.4416	99.3775	0.4856	98.8337
Rank Search	10-Fold Cross-Validation	Fuzzy NN	95.0648	92.0808	96.2335	94.1114	2.3373	89.9286
		Fuzzy Rough NN	**99.6594**	**99.5156**	**99.718**	**99.6167**	**0.245**	**99.0285**
		Fuzzy Ownership NN	99.634	99.6026	99.6536	99.6281	0.3016	99.3018
		VQNN	99.4546	99.3399	99.4876	99.4138	0.4455	98.8973
		OWANN	99.4372	99.3314	99.4586	99.3963	0.4707	98.8638

Here, it is observed that Fuzzy-Rough nearest neighbour classification technique with rank search feature selection method provides better accuracy and low false alarm rate compared to other classification techniques.

On analysing the performance of different classifiers in combination with different ranking and search methods, it is found that Fuzzy-Rough nearest neighbour classification technique with rank search method performs much better in comparison to all other combinations.

Further, we have compared our results with some of the important results reported by other researchers, which is presented in Table 8. It is observed that there is significant improvement in terms of detection rate and false alarm rate. This shows the efficacy of our approach.

Table 8 Comparison of results between the proposed approach with that of the existing ones

Author	Dataset	Feature Selection Method	Classifier Techniques	Detection Rate	False Alarm Rate
Li et al.(2007) [22]	KDD Cup 99	Chi Squared Attribute Evaluator	Transductive Confidence Machines for K-Nearest Neighbour {TCM-KNN)	99.6%	0.1%
Kavitha et al. (2012) [23]	KDD Cup 99	Best First Search	Fuzzy Rule based Intrusion Detection (FRID)	95.47%	10.63%
			Intuitionistic Fuzzy Rule based Intrusion detection (IFRID)	97.86%	5.03%
			Emerging Neutrosophic Logic Classifier Rule based Intrusion Detection (ENLCID)	99.02%	3.19%
Chen et al. (2009) [24]	KDD Cup 99	Rough Set	Support Vector Machine (SVM)	86.72%	13.27%
Sindhu et al. (2012) [25]	KDD Cup 99	Wrapper Approach	Neurotree	98.38%	Not Provided
Sadek et al. (2013) [26]	NSL-KDD	Rough Set	Neural Network with Indicator Variable (NNIV)	96.7%	3.0%
Our Hybrid Approach	**NSL-KDD**	**Greedy Stepwise Search**	**Fuzzy Ownership NN**	**99.6145%**	**0.309%**

7 Conclusions

Building effective intrusion detection models is a challenging task. One of the approaches widely tried out is to classify user behaviour and raise alarms on detecting any anomalous behaviour. Keeping this in view several classifiers have been used but none of the classifier alone is capable of producing acceptable performance. Therefore, work has begun to design hybrid classifiers to improve upon the performance of IDS. The present research is a step forward in this direction where a hybrid model has been proposed with the help of five classifiers and two different categories of feature selection methods. The performances of the classifiers have been evaluated on the basis of accuracy, detection rate, false alarm rate, fitness value, etc. It is observed that the Fuzzy-Rough nearest neighbour classification technique with rank search method performs better in terms of detection rate and reduced false alarms than its counterparts. This observation can certainly help IDS developers in achieving greater accuracy and reducing false alarms. In future, we shall explore application of other hybrid approaches to further improve upon the detection rate and even classify specific attack types.

REFERENCES

[1] Abraham A, Thomas JP, Chebrolu S (2005) Feature deduction and ensemble design of intrusion detection systems. Computers & Security 295-307. doi: 10.1016/j.cose.2004.09.008

[2] Dubois D, Prade H (1992) Putting rough sets and fuzzy sets together. In: Huang S (ed) Intelligent Decision Support, Springer, Netherlands, pp.203-232 .

[3] Pawlak Z (1991) Rough sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishing. Springer, Netherlands

[4] Zadeh L (1965) Fuzzy sets. Information and Control 338-353. doi: 10.1016/S0019-9958(65)90241-X.

[5] Gong, S. (2011) Feature Selection Method for Network Intrusion Based on GQPSO Attribute Reduction, International Conference on Multimedia Technology (ICMT), 6365 – 6368. doi: 10.1109/ICMT.2011.6003117

[6] Hoque MS, Mukit MA, Bikas MAN (2012) An Implementation of Intrusion Detection System using Genetic Algorithm, International Journal of Network Security and Its Applications (IJNSA), .109-120. doi: 10.5121/ijnsa.2012.4208

[7] Zhou YP, Fang JA. (2009) Intrusion Detection Model Based on Hierarchical Fuzzy Inference System". Second IEEE International Conference on Information and Computing Science. IEEE Computer Society, 144–147. doi: 10.1109/ICIC.2009.145

[8] Tong X, Wang Z, Yu H (2009) A research using hybrid RBF/Elman neural network for intrusion detection system secure model. Computer Physics Communications 1795-1801. doi: 10.1016/j.cpc.2009.05.004

[9] Mohamadi H, Habibi J, Abadeh MS (2008) Misuse intrusion detection using a Fuzzy-Meta-heuristic approach. In Proceedings of 2nd Asia Intl. Conference on modeling and simulation 439-444. doi: 10.1109/AMS.2008.128

[10] Panigrahi **A,** Patra, MR (2015) An Evolutionary Computation based Classification Model for Network Intrusion Detection, International Conference on Distributed Computing and Internet Technology (ICDCIT-2015) 318-324. doi: 10.1007/978-3-319-14977-6-31

[11] Killer JM, Gray MR, Givens JA (1985) A Fuzzy K-Nearest Neighbour Algorithm. Systems Man and Cybernet 580-585. doi: 10.1109/TSMC.1985.6313426

[12] Jesen R, Cornelis, C (2008) A New Approach to Fuzzy-Rough Nearest Neighbour Classification. Rough sets and current trends of computing 310-319. doi: 10.1007/978-3-540-88425-5_32

[13] Yager, RR (1988) On ordered weighted averaging aggregation operators in multicriteria decision making, Systems, Man and Cybernetics 183-190. doi: 10.1109/21.87068

[14] Tavallaee M, Bagheri E, Lu W, Ghorbani A (2009) A detailed analysis of the KDD CUP 99 data set. Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defence Applications 1-6. doi: 10.1109/CISDA.2009.5356528

[15] Kohavi R, John GH (1997) Wrapper for feature subset selection . Artificial Intelligence 273-324. doi: 10.1016/S004-3702(97)00043-X

[16] Koller D, Sahami M (1995) Toward optimal feature selection. In Proceeding of International Conference on Machine Learning.284-92.

[17] Han J, Kamber M (2006) Data Mining Concepts and Techniques, 2nd edn, Morgan Kaufmann, San Francisco.

[18] Holte RC (1993) Very simple classification rules perform well on most commonly used datasets. Machine Learning 63-90. doi: 10.1023/A:1022631118932

[19] Kononenko I (1994) Estimating attributes: Analysis and extension of relief. In Proceedings of the Seventh European Conference on Machine Learning. 171-182. doi: 10.1007/3-540-57868-4_57

[20] Marko RS, Igor K (2003) Theoretical and empirical analysis of relief and relief. Machine Learning Journal 23-69. doi: 10.1023/A:1025667309714.

[21] Rich E, Knight K (1991) Artificial Intelligence. McGraw-Hill, 2nd Edition, New-York

[22] Li Y, Guo L (2007) An Active Learning based TCM-KNN Algorithm for Supervised Network Intrusion Detection. Computers & Security. 459-467. doi: 10.1016/j.cose.2007.10.002

[23] Kavitha B, Karthikeyan S, Maybell PS (2012) An Ensemble Design of Intrusion Detection System for Handling Uncertainty using Neutrosophic Logic Classifier. Knowledge-Based Systems 88-96. doi: 10.1016/l.knosys.2011.12.004

[24] Chen RC, Cheng KF, Hsieh CF (2009) Using Rough Set and Support Vector Machine for Network Intrusion Detection. International Journal of Network Security and its Applications (IJNSA) 1-13.

[25] Sindhu SSS, Geetha S, Kannan A (2012) Decision Tree based Light Weight Intrusion Detection using a Wrapper Approach. Expert System with Applications 129-141. doi: 10.1016/j.eswa.2011.06.013.

[26] Rowayda A, Sadek M, Soliman S, Elsayed HS (2013) Effective Anomaly Intrusion Detection System based on Neural Network with Indicator Variable and Rough set Reduction. International Journal of Computer Science Issues (IJCSI) 227-233.

SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM

TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

ISSN 2054-7390

Design of a Voice Based Intelligent Prototype Model for Automatic Control of Multiple Home Appliances

Akanbi C.O. and Oladeji D.O.

Department of Information and Communication Technology

Osun State University, Osogbo, Osun State Nigeria.

akanbico@uniosun.edu.ng

ABSTRACT

The revolution in Information Technology(IT) and Artificial Intelligent has provided innovations in diverse kinds of home automation where appliances can be controlled effortlessly and seamlessly .In this paper a review of several home automation technologies was presented and the advantages voice based technology was considered trending among other technologies in terms of flexibility and cost and comfort. In this paper, the conceptual prototype design model and the circuit diagram for controlling multiple home appliances at the same time were presented. The model uses Arduino Uno microcontroller system to control multiple home electrical or electronic appliances with maximum of 5 volts. The design transmits voice command through user wireless microphone (attached to the user cloth) to the intelligence micro controller via graphical user interphase (GUI). The control software developed for the prototype model was implemented using C# programming language, Microsoft Visual Studio.Net and Microsoft Speech recognition system.

Keywords: Home Automation , Artificial Intelligence, Arduino Uno micro controller, Prototype Model

1 Introduction

The revolution in Information Technology (IT) and Artificial Intelligent has provided innovations in diverse kinds of home automation where appliances can be controlled effortlessly. Home Automation Technologies, is the controlling and monitoring of home appliances in a unified system. These include lighting, heating, and even home electronics. This automation systems range from simple remote control of lighting to complex computer/micro-controller based networks with varying degrees of intelligence and automation. Home automation is adopted for reasons of ease of control for all categories of users, security and energy efficiency by the use microprocessor-based intelligence to integrate or control electronic and electrical devices and systems in the homes such as fans, lights, air conditioners, television sets, security cameras, electronic doors, computer systems, audio/visual equipment etc [1].

Several technologies have been considered in literature for implementation of home automation systems These includes : Power-line Carrier System, land line Telephone system (analog telephone service), infrared light, Bluetooth technology, Short Message Service(SMS),Web method and voice controlled system for implementation ([2], [3], [4], [5], [6], [7].

DOI: 10.14738/tmlai.42.1899
Publication Date: 30th Mar, 2016
URL: http://dx.doi.org/10.14738/tmlai.42.1899

Whichever technology adopted, Users prefer ubiquitous access to home devices instead of been uncomfortably forced to go physically to the nearest control points. Voice control automation technologies which is a trending issue has a better advantages in this regard over others technologies in terms of flexibility, easy access to the device and security.

In this paper, a prototype model of a voice based micro controller system for a home automation system is presented for controlling electrical or electronic appliances. The design transmits voice command through user wireless microphone (attached to the user cloth) to the intelligence micro controller via graphical user interphase (GUI). This provides a comfortable, efficient and cost effective model for all categories of users such as abled, disabled and elderly people in a home to control electrical and electronic gadgets such as fans, lights,etc.[7], [8].

2 Literature Review

[2] presents a smart home wireless remote control device that permits elderly people with physical challenges, in particular, handicapped and disabled people, to command their desired devices without moving around to the nearest control point. The technology uses XBee communication transceivers that receives wireless command signals and activates appliances by triggering the associated electronic relays to achieve the ON/OFF functionality. Also [9] proposed a home appliance control system that uses Infrared ray and power line communication to control the home appliances system. This system helps user to checks the status of appliances and controls them remotely from everywhere through their cellular phone or Internet.

[10] present SMS based wireless home appliance control system (HACS) for automating appliances and security. The proposed a system for controlling home appliances remotely that is useful for the people who are not at home mostly. The system provides security and controls the home appliances such as AC, lights and alarms. The system is implemented by SMS technology to transfer data from sender to receiver over GSM network. One or more computers can be used to control the home appliances. System send an alert SMS to authorized user when any intrusion is detected and user can in turn respond in order to overcome the situation. Moreover user can send SMS to system to get the status of home appliances and controlling them.

The design of voices/speech-based applications is dominated by the underlying Asynchronous Speech Recognition (ASR) technology. From a technological perspective, two broad types of ASR exist according to [11] which are Direct Voice Input (DVI) and Large Vocabulary Continuous Speech Recognition (LVCSR). DVI devices are primarily aimed at voice command-and-control, typically configured for small to medium sized vocabularies (up to several thousand words). It usually required responding immediately to a voice command employs word or phrasing spotting techniques.

[12] proposed a home automation system based on voice recognition. The system implements Automatic Speech Recognition using speech processor and MATLAB coding .The prototype developed can control electrical devices in a home or office. [13] proposed speech recognized automation system using speaker identification through wireless communication. The prototype developed can control electrical devices in a home or office wirelessly. The system implements ASR (Automatic Speech Recognition) using speech processor and speaker identification through MATLAB coding and MFCC algorithm.

[14] proposed a voice controlled wireless smart home system for elderly and disabled people. The proposed system has two main components namely voice recognition system, and Zig Bee wireless system. Lab View software was used to implement the voice recognition system. Based on the received data at the wireless receiver associated with the appliances desired switching operations are performed.

[15] proposed a model for the physically challenged and elderly people . The study demonstrates a system that can be integrated as a single portable unit and allows one to widely control light, fans, air conditioners, television sets, security cameras, electronic doors etc. The system is controlled from a microphone, FLEX sensor and PIC. The chip sends the voice command in binary sequence to the microcontroller. The base station unit takes the decision and sends the commands to the remote station by Zigbee transceiver. The sensor unit is capable of detecting when the user enters or leaves the room by measuring the change in signal strength between the access points and can accordingly turn on/off appliances such as lights and fans and in the meantime send its status back to the base station. XBee transceivers are used to eliminate the need for large amount of wiring between the processor and the appliances

While most of the literature discussed above use Zigbee transceiver and other technologies that are very expensive technology to model and implement, The design model presented in this paper uses cheaper and reliable technology - Arduino Uno IDE Micro Controller to control home appliances .

3 System Design

The design consideration in this paper is modeling a voice automation system that allows user to be anywhere in a building and remotely control electrical and electronic gadgets. The system conceptual design is presented in figure 1 consisting of four major modules; the Sensor (wireless microphone), Speech Recognising Synthesizer (SRS)/Graphical User Interphase (GUI) (in a laptop/desktop), Arduino Uno micro controller.

Wireless microphone (as sensor) captures the user's voice command. This command is transmitted to the Arduino Uno microcontroller via the GUI software developed. The users (abled, disabled and elderly people) give command through the wireless microphone connected to the PC (wirelessly), the GUI Software developed listens for audio command and transmits its processed output to the Arduino Uno micro controller. The GUI software program serves more or less as a filter since it recognizes only speech that is found in the grammar library.

A special library header (using system synthesizer) is used for this technology using C# programming language codes. A laptop or desktop connected via COM4 to an Arduino serves as the controller of the system.

The Arduino Uno micro controller was connected and powered through a 5V power supply via a USB cable from the computer. The micro controller has pins on it and a light indicator that shows when it is been powered ON. Each pin outputs 5V power supply to a bread board via a jumper wire. Another jumper wire takes out 5v from the bread board to a relay switch connected to a Vero board. Two relay switches are used to switch light bulb and Fan to ON and OFF states. The command from the user is sent to the Arduino Uno Micro Controller via COM 4, the light comes on/off.

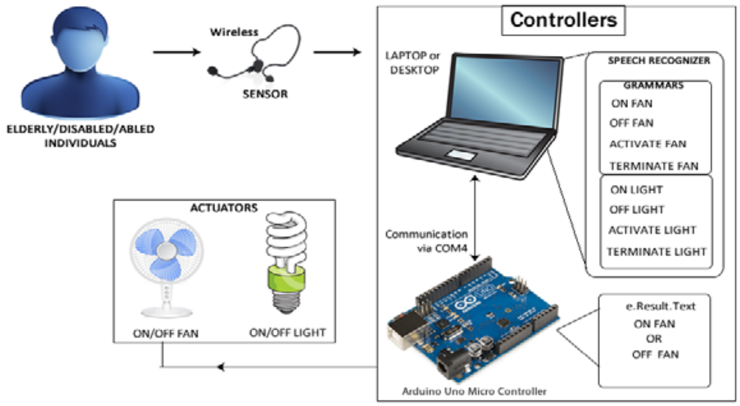


Figure 1: Conceptual Model for Voice Control System

This research work switches a light bulb or fan ON and OFF so eight (8) pre-defined grammars are used, which are.

a) Switch ON light.
b) Switch OFF light.
c) Activate light.
d) Deactivate light.
e) Switch ON fan.
f) Switch OFF fan.
g) Activate fan.
h) Deactivate fan.

As voice signal is transmitted from the microphone to the PC, the system synthesizer deciphers and puts into appropriate tokens whatever has been detected and forwards it to the Arduino, a match between this and the library pre-programmed into the Arduino gives a positive for whatever command the user desires. In this case either to ON/OFF/ ACTIVATE/TERMINATE a light bulb, ON/OFF/ ACTIVATE/TERMINATE a fan.

4 Implementation

For the implementation, four modules were developed – Control software module, Micro controller firmware programmable system module, Connectivity module (using jumper wires) and relay module to power AC/DC electric fan. The control software programmable module was written using C# programming language which is deployable on any computer running Microsoft Windows operating system. The control software communicates with the micro controller through COM4 and serial port 1.

The Graphical User Interface (GUI) was developed using Microsoft Visual Studio.Net. The code was written in a code editor frame. Since the Arduino micro controller is a serial communication device, therefore the GUI voice control / activation system application must communicate with the Arduino board via a serial communication channel. The physical connection to the Arduino micro controller is a USB cable, hence 'COM' serial communication channel was used. From the Arduino IDE, COM4 was selected, therefore all message passed/transported to the Arduino Uno micro controller must pass through COM4 communication pathway via USB cable.

The control module encompasses a grammar library that holds the acceptable commands from the users. Also, the Microsoft speech recognition engine or speech synthesizer engine is used in the application to enable the control software application decipher the speech spoken. When a voice command is spoken, the software matches it with what is in the grammar library. If a match is found, it then uses an 'if' control structure to act on the necessary command for the micro-controller to respond appropriately.

The circuit diagram is shown in figure 2 below. In the diagram, four outputs are sent from the control software to the micro-controller, 1, or 0 for the Bulb while 'A' or 'a', are for the fan. The micro-controller is programmed to accept all four inputs through COM4 and serial port 1 under the setup control structure of the program, it then sends out a voltage signal which is 5V via pin 6 or pin 1. A jumper wire was used to send the 5V from the micro controller to the bread board. Also, from the from the bread board, another jumper wire was used to send the 5V to a pin called 'IN' on the relay module. The relay module has three pins which are 'IN', 'GND', &'VCC', the 5V goes to the IN on the relay

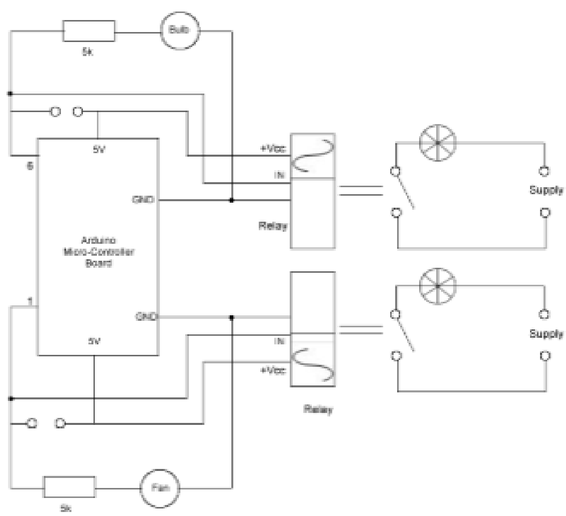


Figure 2: Circuit Diagram of the Voice Controlled Prototype Model

module, and the VCC is connected via jumper wire to the VCC on the micro controller, and the GND is connected to the GND on the micro controller. The relay module also has three holes that accept connection from an AC/DC load which are NO, COM and NC. The live wire which is the electrical wire is connected to the COM and NO for a circuit to be created. When the key closes inside the relay module due to the voice command spoken and interpretation by the micro controller, a complete circuit is created and the electric fan or light goes to ON state. Figure 3 shows a picture of the implementation setup of the Arduino Uno and the breadboard connected with jumpers.

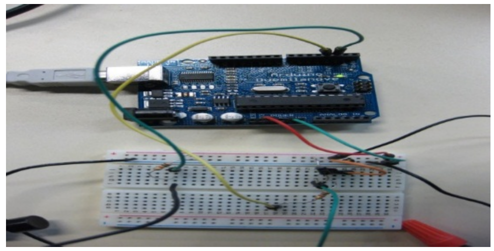


Figure 3: Arduino Uno Micro controller connected to Bread Board

4.1 Switching on the light or fan

When the serial Port method is called and an argument is passed onto it like serial Port ("1"), this tells the Arduino Uno micro controller via COM4 that is should supply 5V power supply to pin6. Then pin6 supplies the bread board through looping of a jumper wire from the pin6 on the micro controller to the Vero board. A relay switch is connected / mounted on the Vero board. This 5v coming from the Vero board triggers an electromagnetic field inside the relay switch to close the switch for a complete circuit to take place. When the complete circuit takes place, the light bulb goes ON. The same principle applies for turning on the fan however, to turn on the fan the argument to the serial port is ("A"). The micro program code snippet (written in C language) for Arduino Uno Controller for switching the bulb on and off is shown in figure 4 below .

```
int message = 0; //this holds one byte of the serial message

intBULBPin = 6; //this is the pin that the jumper wire is connected to
intBULB = 0 //the value or brightness of the LED (off state 0v).
void setup(){
Serial.bigin(9600); //set serial to 9600 baud rate
If(serial.available()>0) //check to see if there is a message
{
Message = Serial.read(); // Put the serial input into the message
If (message == '1'){ //If capital A is received
BULB = 255; //set BULB to 255 (ON) state of 5v
Serial.println("BULB ON"); //send back BULB ON}
if(message == '0'){ // if lowercase a is received
BULB = 0; // set BULB to OFF state of 0v
Serial.println("BULB OFF"); //Send back BULB OFF
// end of code snippet}}}
```

Figure 4: C language code snippet for the Arduino Uno Microcontroller software to light the bulb

To activate the system for speech / voice authentication to take place, Jarvis plus the command syntax in the grammar library i.e. Jarvis + ON fan. Conversely Jarvis + OFF deactivates the system.

4.2 Switching off the light or fan

When the serial Port method is called and an argument is passed onto it like serial Port ("0") or serial Port("a"), this tells the Arduino Uno micro controller via COM4 that is should supply 0v power supply to pin6 or pin1. Then pin1 supplies the bread board through looping of a jumper wire from the pin1 on the micro controller to the Vero board. The light bulb is connected to the Vero board as an indicator that 0v supply is in action. A relay switch is connected / mounted on the Vero board. Since a relay needs a little supply of 5V to trigger the switch inside to complete the circuit and cannot be found in 0v then the switch key opens and an incomplete circuit is created and the fan or the light is OFF.

5 Conclusion

This paper has presented a model of a voice based controlled system for home appliances. This has effortlessly, solved the problem of movement for all categories of people such as elderly, disabled or even abled people in an attempt to control home appliances. Though this work is limited to home appliances with maximum of five volt (5V), further works can accomodate home appliances with above five volts (5V+). Also, we intend to integrate variable control functions to improve the system versatility such as providing control commands other than ON/OFF commands. For example "Increase Temperature", "Dim Lights" etc. Another area to consider is controlling of home devices via web and mobile platform.

REFERENCES

[1] Karimaa A. (2011) Mobile and Wireless Access in Video Surveillance System International Journal of Digital Information and Wireless Communications (IJDIWC) 1(1): 267-272

[2] Bilal G. and Khaled A. (2015) Smart Home Automation System for Elderly, and Handicapped People using XBee "International Journal of Smart Home" Vol. 9, No. 4

[3] Hammed B. (2012) "Design & Implementation of Smart House Control Using LabView". International Journal of Soft Computing and Engineering, January, Vol. 1, Issue 6, pp. 98-106.

[4] Mazidi M. Mckinlay R., and Causey D(2007), "PIC microcontroller and embedded systems", 1st ed, Prentice Hall

[5] Dutta K., Rai P, and Shekhe V(2012), "Microcontroller based voice activated wireless automation system".International journal of electrical, electronics & communication engineering, (2012), Vol. 2, No. 8, pp.642-649.

[6] Sriskanthan N. and Karand T.(2002),"Bluetooth Based Home Automation System", Journal of Microprocessors and Microsystems, Vol. 26, pp.281 289, Elsevier Science

[7] Shepherd R.(2001, "Bluetooth wireless technology in the home" , Journal of Electronics and Communication Engineering, Vol. 13, no. 5, pp. 195

[8] Alkar A.Z, and Buhur U., "An Internet Based Wireless Home Automation System for Multifunctional Devices," IEEE Transactions on Consumer Electronics, vol. 51, no. 4, pp. 1169-1174, Nov. 2005.

[9] Nguyen T.M, Lee D.G., Seol Y.H. Yu M.H, Choi D. (2007), "Ubiquitous Access to Home Appliance Control System using Infrared Ray and Power Line Communication", ICI 2007,3rd IEEE/IFIP International Conference in Central Asia, Tashkent, Uzbekistan, vol 1, pp1-4,26-28 .

[10] Khiyal M., Khan A. and Shehzadi E. (2009), "SMS based wireless home appliance control system (HACS) for automating appliances and security". Issues in Informing Science and Information Technology, Vol. 6, pp. 887-894.

[11] Ananya. M.S and Sakthivel.S. (2015): A HMM Based Automatic Speech Recognition System to Generate Software Requirement Specification. International Journal For Trends In Engineering & Technology Volume 7 Issue 1

[12] Prabhakar V. M. and Amol C.W.(2013) "Speaker Identification Based Automation System Through Speech Recognition" in International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 1

[13] Suralkar S. R., Amol C.W. and Prabhakar V. M., "Speech Recognized Automation System Using Speaker Identification through Wireless Communication" IOSR Journal of Electronics and Communication Engineering, Volume 6, Issue 1,2013, pp 11

[14] Thoraya O. Haliemah R; Ali A.; Muhammad R.and Mussab M. (2014):International Journal of Wireless & Mobile Networks; Vol. 6 Issue 1, p47

[15] Parameshachari B D, Sawan K. G.,Gooneshwaree H.and Tulsirai T.G.(2013): A Study on Smart Home Control System through Speech. International Journal of Computer Applications Volume 69, No.19.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

ISSN 2054-7390

An Efficient Brain Tumour Extraction in MR Images using Ford-Fulkerson Algorithm

C. Latha and K. Perumal
Department of Computer Application, Madurai Kamaraj University
rocela16@gmail.com; perumalmkucs@gmail.com

ABSTRACT

Brain tumor division intends to discrete the diverse tumor tissues, for example, dynamic cells, necrotic core,and edema from typical cerebrum tissues of White Matter (WM), Gray Matter (GM), and Cerebrospinal Fluid (CSF). MRI-basedbrain tumor division studies are drawing in more consideration lately because of non-invasiveimaging and great delicate tissue difference of Magnetic Resonance Imaging (MRI) pictures. With the improvement of very nearly two decades, the inventive methodologies applying PC supported strategies for the sectioning brain tumor are turning out to be more develop and coming closer to routine clinical applications. The reason for this exploration work is to give a far-reaching review to MRI-based brain tumor division techniques. To consider and characterize the tumor pictures, a couple of well-known Edge Detection Techniques have been proposed as of late. This examination work has distinguished KWT (K-Means, Watershed, and Texture) Segmentation Technique and executed and contemplated. From our test comes about, this examination work uncovered that this model neglects to make productive groups order force causes poor tumor characterization precision. This is one of the real issues to anticipate the tumor design and to address this issue, the Ford-Fulkerson Segmentation Technique is proposed and concentrated altogether as far as Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Classification Accuracy. Results set up that the proposed Ford-Fulkerson Segmentation Technique beats KWT in terms of Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Classification Accuracy.

Keywords: K-Means, Watershed, Texture, Edge Detection, Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Classification Accuracy.

1 Introduction

Image Processing [1, 2, 3] is being utilized as a part of numerous continuous applications and medicinal applications also. Some imperative ranges are Computer Vision for errands like the route of Robots, medicinal field, for example, Disease Diagnosis from CT output and MRI Images, Remote Sensing, Video Processing and ID of number plates of moving vehicles and so forth. The reason for Digital Image Processing is robotizing diverse errands and image division is an unavoidable stride in it. Image Segmentation alludes to apportioning a picture into a few sections, taking into account Textures, Colour, Edges or Regions in the picture. It goes for isolating the picture into outwardly homogeneous and particular areas. Shading is viewed as an applicable component when managing the impression of static

DOI: 10.14738/tmlai.42.1875
Publication Date: 23rd April, 2016
URL: http://dx.doi.org/10.14738/tmlai.42.1875

and moving pictures. Visual contrast [3,4] is useful to filter information present in each colour component and to distinguish among alike gray-scale intensities. The mixture of colour and texture has been proved to achieve better results and could be exploited more successful.

Image Segmentation is a critical part of Digital Image Processing. The Image Segmentation can give an outcome that streamlines the presentation of an Image and makes the picture investigation simpler. In the division, a name is doled out to each pixel that is having comparable qualities, similar to shading, surface or force, which will help to partitioned the locales and distinguish the items and their limits. In any case, the issue while handling is the possibility of over-segmentation or under-segmentation.

2 Related Work

Many image segmentation calculations are generally utilized as a part of a few regions. Such segmentation systems may be comprehensively grouped into intermittence based division and comparability based segmentation [1, 2, 3, 4, 5]. To start with class incorporates picture division calculations like edge identification by identifying the edges or pixels between distinctive districts that have quick move in power are removed and recent incorporates division calculations, for example, locale developing and area part and combining or thresholding system. Image Segmentation has an imperative part in therapeutic imaging for the most part in diagnosing variations from the norm in pictures of human body parts. The method for dividing image fluctuates from picture to image furthermore relies on upon the reason for segmentation.

K-Means Clustering Methods [1, 2, 3] has been utilized for therapeutic image segmentation particularly in MR Images of the human mind. For CT sweep pictures, some dynamic shape models [4, 5] can be utilized and various model methodologies [4] are there for sectioning ultrasound pictures of the heart.

Therapeutic field utilizes distinctive division techniques, [4, 5] shows the utilization of segmentation of punctuating examples in fluorescence microscopic images. It's factual displaying helps the different covers meet from an arbitrary starting setup to an applicable one. Another work in [3,5] presents a system to fragment countless from 3-D pictures portrayed by non-homogeneous power and angle sign and skilled to finish surface discontinuities with no bargain in the middle of accuracy and capacity to coordinate the deficient shapes. The division system is a summed up variant of the Subjective Surfaces procedure. But most challenging segmentations in medicinal imaging consist of the processing of edgeless images such as histology images which consists of bright field microscopy images of haematoxylin and eosin (H&E)-stained slices of tissues. Some segmentation works on histology dataset [3, 4, 5], which deals with slices of tera-toma tumour [4, 5], where the intensity neighbourhoods are used to segment bone, cartilage, and obese. Another method for identifying the histological grade of breast cancer based on model classification and image analysis algorithm is described clearly in [3].

From the literature survey[1,2,3,4,5], it was seen that a couple of famous Edge Detection Techniques in particular Threshold-based routines, Region-based systems, Classification and Clustering strategies like K-Means, Fuzzy C-Means (FCM), Markov Random Fields (MRF), Bayes, Artificial Neural Networks (ANN), Support Vector Machines (SVM) and a couple of other Segmentation Techniques, for example, Watershed, Ford-Fulkerson, have been proposed.

This research work has recognized two well-known Edge Detection Techniques, in particular, i. KWT i.e. joining K-Means, Watershed and Texture Segmentation Technique and ii. Portage Fulkerson

Segmentation Technique. The exhibitions of these two procedures have been concentrated completely by actualizing them and the execution of Ford-Fulkerson Technique is enhanced by proposing MRF and CRF based Ford-Fulkerson Technique. The current KWT Technique and proposed model are talked about at the accompanying areas.

3 KWT Edge Strength Merging Technique

Recognizing and Eliminating False Edges amid Image Segmentation is one of the testing issues in Image Segmentation Process. The creators Gullanar and et. al. dissected the execution of consolidating K-Means, Watershed and Texture (KWT) Segmentation Technique[3,4]. The ideas and models of KWT are described beneath.

3.1 K-Means Clustering Technique

Among the hard clustering, K-Means clustering [1,3,5] is forever analyst's first decision in view of its straightforwardness and elite capacity. Here, K is the quantities of groups to be indicated. The formal steps included in this calculation are:

1) Randomly choose 'c' cluster centres.

2) Calculate the distance between every data point and cluster centres.

3) Allocate the data point to the cluster centre whose distance from the cluster centre is minimum of all the cluster centres.

4) Recalculate the new cluster centre using:

$$v_i = \frac{1}{c_i}\sum_{j=1}^{c_i} x_j$$

where, 'c_i' denotes the number of data points in i^{th} cluster.

5) Recalculate the distance between every data point and new attained cluster centres.

6) If no data point was reselected then stop, otherwise repeat from step 3).

The goal of the K-Means algorithm is to diminish the squared error function [3]:

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\|x_i^{(j)} - C_j\right\|^2 \tag{1}$$

Here, $\left\|x_i^{(j)} - C_j\right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centrec_j isan indicator of the distance of the n data points from their respective cluster centres.

3.2 Watershed Segmentation

A Watershed [2] definition for the unremitting case can be based on reserve functions. Dependingon the reserve function used one may arrive at diverse definitions. The image f is an element of the space C(D) of real doubleendlessly differentiable functions on a connected domain D with only isolated critical points. Then the geographical distance between point's p and q in D is defined [1] by

$$T_f(p,q) = \inf_{\gamma} \int \gamma \left\|\nabla f(\gamma(s))\right\| ds \tag{2}$$

Where the infimum is over all paths (smooth curves) γ inside D with γ(0) = p, γ(1) = the topographical distance between a point p € D and a set A ⊆ D is de_ned as $T_f(p, A) = MIN_a \in AT_f(p, a)$.$T_f$is theshortest path between p and q with steepest slope.

3.2.1 Watershed Tree Construction

In the multi-scale representation, the tree's leaves speak to the starting allotment of the image area. Inside Nodes speak to locales acquired by combining the areas relating to their kids. The Root Node speaks to the whole picture support. Amongthese lines, the tree speaks to an arrangement of locales at diverse scales, and it can be viewed as a progressive, area based representation of the data picture. Unmistakably, the tree does not encode all conceivable outcomes for blending areas fitting in with the introductory segment, yet just the most valuable combining steps. In this manner, both the combining request and the area model whereupon the tree development procedure depends on must be deliberately picked.

Compute Gradient-Magnitude Image E of F;

Apply the Watershed Transform on E to get initial partition P;

Compute region model of all regions of P;

Assign regions of P to leaves of T

Compute Region Adjacency Graph (RAG) G of P;

While vertices(G) > 1 do

{

Compute edge costs;

Construct Minimum Spanning Tree(MST) Tm on G

Apply Watershed transform on Tm to build new partition P;

Compute region model of all regions of P;

Establish child-parent relations between regions in T at the previous level and regions in P

Compute RAG G of the new partition.

}

End while

3.3 Texture Segmentation and Statistical Analysis

The statistical approach i.e., an image texture considered as quantitative measure at the regions need to analysis. The mathematical model is discussed at the following section.

3.3.1 Edge Detection

The edge detection is used to determine pixels at a specified region are facilitating to determine the texture complexity.

Give us a chance to consider an image as made out of homogeneously textured areas which appeared in the Fig. 1. It was considered and accepted that the ghostly histograms inside homogeneous areas are steady. The Local ghostly histograms illustrative of every district can be registered from windows inside every area. Give us a chance to consider just the force channel until further notice, which gives the power estimation of every pixel as the channel reaction. At that point, the neighborhood unearthly histogram is identical to the histogram of a nearby window. Under the suspicion of ghostly histogram steadiness inside of the district, the nearby histogram of pixel A can be all around approximated by the weighted whole of agent histograms of two neighbouring locales, where the weights relate to region scope inside of the window and in this way demonstrate which area pixel A has a place with.

Given an image with N pixels and M feature dimensionality o, entirely the feature vectors can be compiled into aM $\times$ N matrix, Y. Assuming that there are L illustrative features, the image typical can be expressed as:

$$Y = Z\beta + \varepsilon \tag{3}$$

Where Z is a M$\times$L matrix whose columns are illustrative features, β is an L$\times$ N matrix's columns are weight vectors, and ε is the model error. The representative feature matrix Z can be calculated from manually nominatedwindows

Within each identical region, and β is then estimated by least squares estimation:

$$\hat{\beta} = (ZTZ)\text{-}1ZTY \tag{4}$$

Segmentation is obtained by examining $\hat{\beta}$ where each pixel is assigned to the segment where the corresponding representative feature has the largest weight.

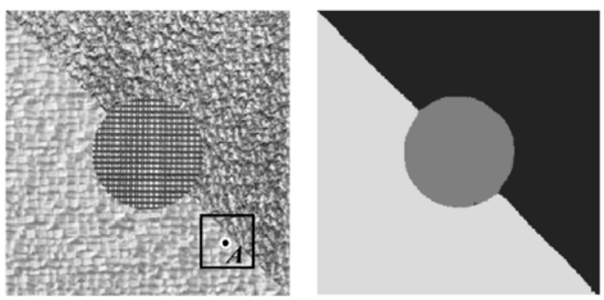


Figure 1: Textured Image with Segmentation

3.4 KWT based Image Segmentation

The KWT [3, 4, 5] is a region-based image segmentation process. This method consists of the following steps. They are

Step 1: Load Image

Step 2: Compute a segmentation function where images' dark regions need to segment

Step 3: Convert a Loaded Image into L* a* b* and compute foreground markers

Step 4: Apply K-Means algorithm and Compute distance and cluster

Step 5: Apply Watershed Segmentation Technique and Texture Segmentation

Step 6: Compute the watershed transform

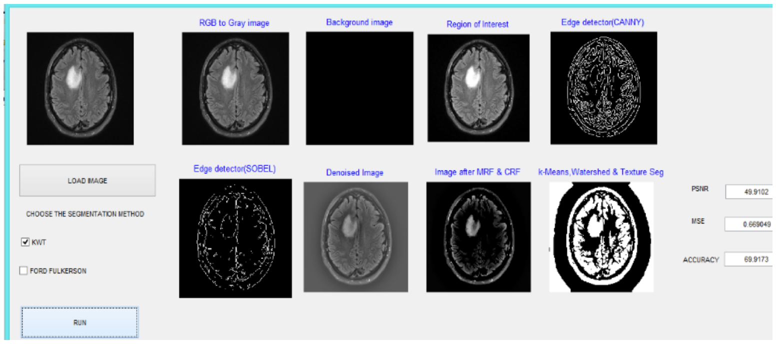


Figure 2: Performance Analysis of KWT

3.4.1 Mean Squared Error(MSE) and Peak Signal to Noise Ratio(PSNR)

The Mean Squared Error (MSE) is the cumulative squared error linking the compressed and the original image and the Peak Signal to Noise Ratio (PSNR) is the measure of Peak Errors. The Statistical model is given below to calculate both the MSE and PSNR.

Mean Squared Error (MSE)

The mathematical model to measure MSE is as given below

$$MSE = \frac{1}{MN}\sum_{Y=1}^{M}\sum_{x=1}^{N}[I(x,y) - I^{'}(x,y]^{2} \quad (5)$$

Where, I(x,y) is the original image, I'(x,y) is its noisy estimated version (which is really the decompressed image) and M and N are the dimensions of the images value for MSE implies lesser error.

Peak Signal to Noise Ratio (PSNR)

The mathematical model to measure PSNR is as given below

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (6)$$

Where, MAXI is the greatest possible pixel value of the image. A higher value of PSNR has always chosen as itsuggests the ratio of Signal to Noise will be developed. The 'signal' here is the original image, and the 'noise' is the errorin reconstruction.

3.5 Identified Problem

From our experimental results, it was noticed that the KWT Image Segmentation Technique i.e., the combining K-Means, Watershed and Texture-based Image Segmentation technique achieved the segmentation result.

However, from our experimental results, this research work revealed that this model fails to make efficient clusters to classify intensity causes poor tumourclassification accuracy. This is one of the major issues to predict the pattern and to address this issue, the Ford-Fulkerson image segmentation Technique is proposed and framework is described at the following Section.

4 FFT: Ford-Fulkerson Technique

As we discussed at the previous section, it was noticed that we need an efficient image segmentation Technique to effectively segment the tumour images to increaseforecast and classification accuracy of Tumour images. The Ford Fulkerson Technique is proposed to address this issue and it is discussed in this Section.

4.1 Clustering Method based on Ford Fulkerson Segmentation Technique

From the literature survey, in fact, it was noticed that most of brain tumour segmentation algorithms are based on classification or clustering methods such as Fuzzy C-Means (FCM), K-Means, Markov Random Fields (MRF), Bayes, Artificial Neural Networks (ANN), and Support Vector Machines (SVM).

In our work, we considered Ford Fulkerson Technique to improve classification and prediction accuracy. This is based on MRF approach.

MRF had been proposed as it can give an approach to incorporate spatial data into the bunching or order process. In grouping strategies, it diminishes both the conceivable issue of covering and the impact of commotion on the outcome. In the specific instance of cerebrum tumour segmentation, if a district is unequivocally named as mind tumour or non-mind tumour, MRF will figure out whether the neighbour of the marked locale is the same. Contingent Random Fields (CRF) had been proposed to construct probabilistic models to section and mark grouping information. MRF and CRF calculations can speak to complex conditions among information sets to pick up a high precision for the results of cerebrum tumour segmentation.

4.1.1 Ford Fulkerson Procedure

The procedure is developed as follows.

Inputs: Graph G with flow capacity C,

s-Source node and *t*-sink node

Output: A flow f from *s* to *t* which is a maximum

Steps:

Step 1: f (u,v) ←0 for all edges (u,v)

Step 2: While there is a path p from *s* to *t* in G_f, such that c_f (u, v) > 0 for all edges (u,v) ∊ p :

{

1. Find c_f(p) = min{c_f (u, v) : (u, v) ∊ p}

2. For each edge (u,v) ∊ p

{

1. f (u, v) ← f (u, v) + c_f (p) (Send flow along the path)

2. f (v,u) ← f(v,u) – c_f (p) (The flow might be "returned" later)

}

}

Step 3: Stop

4.1.2 Flowchart

The flow chart of our proposed Ford-Fulkerson Technique is presented in this section as follows.

Step 1: Load Image

Step 2: Compute a segmentation function where images' dark regions are need to segment

Step 3: Convert a Loaded Image into L* a* b* and compute foreground markers

Step 4: Apply morphological operation and Compute distance and cluster

Step 5: Apply Ford-Fulkerson Image Segmentation Procedure

Step 6: Compute the transform and create final segmented image

5 Results and Discussions

This proposed method is implemented with MATLAB Tool. Here, the number of Clusters K are initialized and values are given as dynamic to make automatic clusters depends upon the available database. The results of the existing and proposed model are given below.

The developed tool calculated both the MSE and PSNR values for the original and final segmented images which are shown in the Figures.

From the segmented image, we need to extract the required information and features. The features are used to predict or classify the Tumours from normal Image.

The experimental results are demonstrated in the Figures Figure. 2, Figure. 3, Figure. 4, Figure. 5, and Figure. 6.

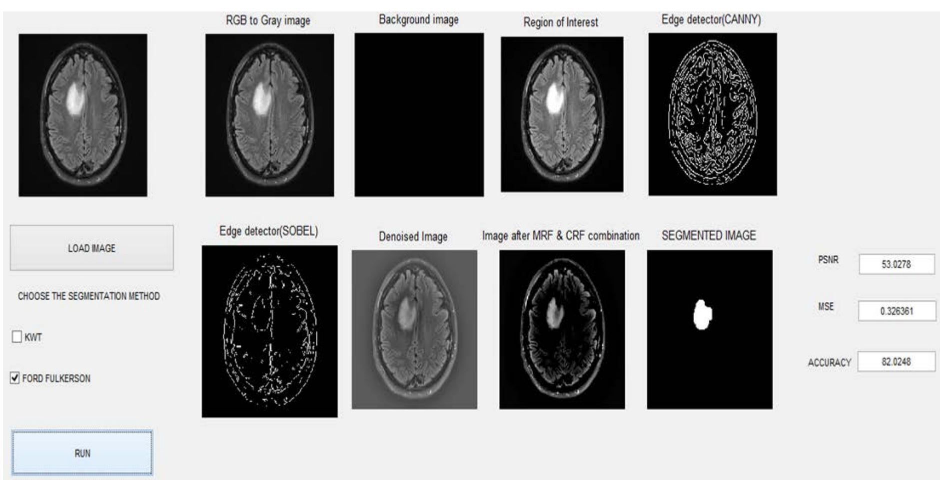


Figure 3: Performance Analysis of Ford Fulkerson

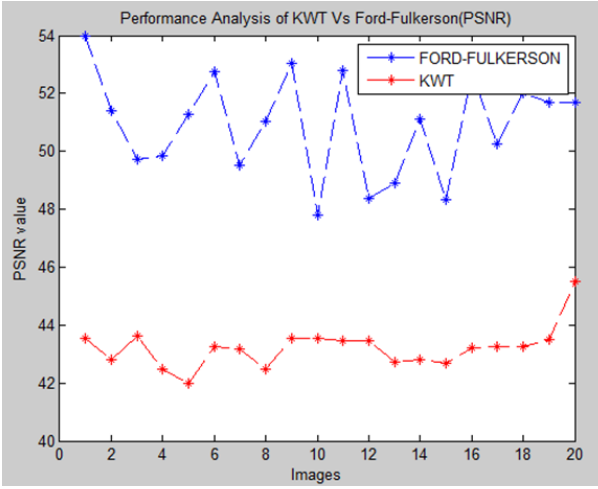


Figure 4: PSNR (KWT vs Ford Fulkerson)

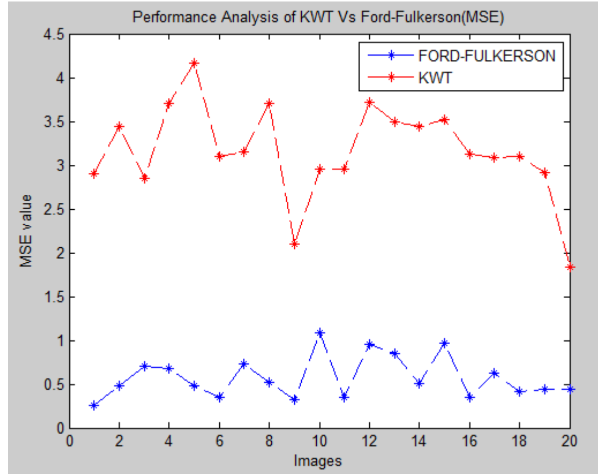


Figure 5. MSE (KWT vs Ford Fulkerson)

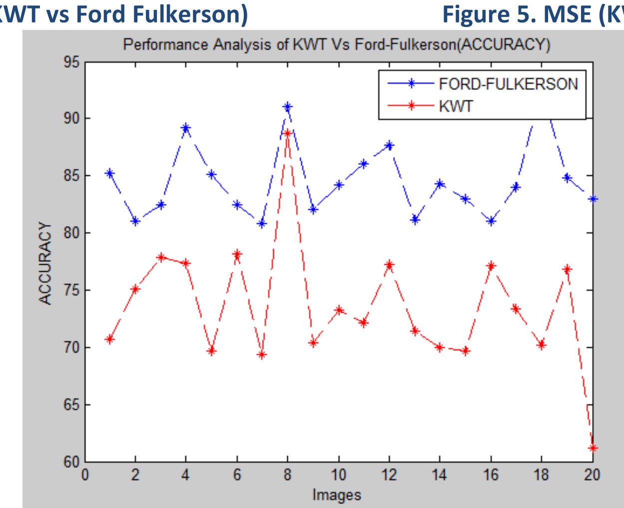


Figure 6: Classification and Prediction Accuracy (KWT vs Ford Fulkerson)

As shown in the Figure. 2 and Figure . 3, the proposed model Ford-Fulkerson was implemented and its performance in terms of PSNR, MSE and Prediction Accuracy for various Tumour Images was identified. From the results, it was noticed that our proposed model is performing well in terms of PSNR, MSE and Prediction Accuracy when compared with existing KWT Segmentation Technique, which are shown in the Figures 4, 5 and 6.

6 Conclusions

This research work has proposed Ford-Fulkerson Segmentation Technique and implemented with MATLABTool. The efficiency of the proposed technique is compared with that of existing KWT (K-Means, Watershed and Texture) Segmentation Technique and established that the proposed model is performing well in terms of Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Classification Accuracy.

REFERENCES

[1] Yong Ning, Xiangjun Zhu, Shanan Zhu, and Yingchun Zhang. 2015. Surface EMG Decomposition Based on K-means Clustering and Convolution Kernel Compensation.

[2] Raffaele Gaetano, Giuseppe Masi, Giovanni Poggi.2015. Marker-Controlled Watershed-Based SegmentationofMultiresolution Remote Sensing Images.

[3] KarteekaPavan, k and SrinivasaRao, ch.2015.A Sequential K-Means Clustering for Mammogram Segmentation.

[4] DibyaJyotiBora, and Anil Kumar Gupta. 2014.A Novel Approach Towards Clustering Based Image Segmentation.

[5] Gullanar, M. Hadi and Nassir H. Salman. 2015. A Study and Analysis of different Edge Detection Techniques.