International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:4, No:12, 2010

Certain Data Dimension Reduction Techniques for application with ANN based MCS for Study of High Energy Shower

Gitanjali Devi, Kandarpa Kumar Sarma, Pranayee Datta, and Anjana Kakoti Mahanta

Abstract—Cosmic showers, from their places of origin in space, after entering earth generate secondary particles called Extensive Air Shower (EAS). Detection and analysis of EAS and similar High Energy Particle Showers involve a plethora of experimental set-ups with certain constraints for which soft-computational tools like Artificial Neural Network (ANN)s can be adopted. The optimality of ANN classifiers can be enhanced further by the use of Multiple Classifier System (MCS) and certain data - dimension reduction techniques. This work describes the performance of certain data dimension reduction techniques like Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Self Organizing Map (SOM) approximators for application with an MCS formed using Multi Layer Perceptron (MLP), Recurrent Neural Network (RNN) and Probabilistic Neural Network (PNN). The data inputs are obtained from an array of detectors placed in a circular arrangement resembling a practical detector grid which have a higher dimension and greater correlation among themselves. The PCA, ICA and SOM blocks reduce the correlation and generate a form suitable for real time practical applications for prediction of primary energy and location of EAS from density values captured using detectors in a circular grid.

Keywords—EAS, Shower, Core, ANN, Location.

I. Introduction

Cosmic showers have extremely high energy and produce complicated processes during their transit through the atmosphere of the earth. Such interactions with atmospheric nuclei result in certain secondary particles called Extensive Air Shower (EAS) [1] [2]. The study of these EAS involves the measurement of the position, size, primary energy, time extent of the events and other related factors. Detection and analysis of EAS comprises of complex measurement and detection equipments with several constraints due to partial knowledge regarding interactions of shower particles and primary energies [2] for which inaccuracies are observed. These difficulties make the analysis of showers a tedious task and requires constant support from expensive experimental set-ups which is a limitation for visualization, conceptualization and monitoring of EAS events. These constraints necessitate soft-computational approaches which can be used to predict primary energy and locations of shower events. Soft computational tools like Artificial Neural Network (ANN)s can be trained to adapt to situations and learn the variations taking place. The knowledge developed using *apriori* references patterns can be utilized for predicting future events as an aid to expand the knowledge of EAS and can be made a part of physical experiential apparatus to facilitate adaptive orientation of monitoring and analysis of EAS phenomena that too in real time.

Several works exist which have used multiple approaches to analyze EASs and thereby develop applications suitable for shower events. A work by D. Hanna [3] reports application of ANNs for EAS. Another work by J C Perrett and J T P M van Stekelenborg [4] describes the implementation of an ANN to estimate the core position and energy of EASs recorded by the South Pole Air Shower Experiment (SPASE) [5]. Another work of similar nature is [6]. This work discusses the possibilities of using ANNs for individual EAS data evaluation. A work as cited in [7] uses ANN for providing a mass likelihood distribution for each measured shower based on its multi-parameter training with simulated showers. Another work by A. Chilingarian . et. al [8] is related to ANN models to recognize the experimental EAS without known primary energy.

Application of soft computational methods to EAS analysis involves the configuration of tools like ANNs as classifiers. The ANN classifiers receive inputs from a host of detectors placed in a circular grid. The detectors are placed in certain assumed regularity and record shower events as they enter the atmosphere of the earth. It means that a shower event can be recorded simultaneously by several detectors and the measured parameters reported for analysis. These data can be highly correlated which at times makes the computational complexity of ANN classifiers increase. The end result is a less optimal decision generated by the ANN. Such a fact is applicable to systems that use Multiple Classifier System (MCS)s for EAS and other similar analysis. This work describes certain data dimension reduction techniques like Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Self Organizing Map (SOM) approximators for application with an MCS formed using Multi Layer Perceptron (MLP), Recurrent Neural Network (RNN) and Probabilistic Neural Network (PNN). The data inputs are obtained from an array of detectors placed in a circular arrangement resembling a practical detector grid which have a higher dimension and greater correlation among themselves. The PCA, ICA and SOM blocks reduce the correlation and generate a form suitable for real time practical applications for prediction of

Gitanjali Devi is with the Deptt. of Physics, Lalit Chandra Bharali College, Guwahati - 781011, Assam, India. e-mail: (gitanjalilcb@gmail.com); Kandarpa Kumar Sarma and Pranayee Datta are with Deptt. of Electronics and Communication Technology, Gauhati University, Guwahati - 781014, Assam, India e-mail: (kandarpaks@gmail.com)

Anjana Kakoti Mahanta is with Deptt. of Computer Science, Gauhati University, Guwahati -781014, Assam, India.

primary energy and location of EAS from density values captured using detectors in a circular grid. The objective is to formulate optimal sets of density values of EAS events such that subsequent tasks of prediction or detection is accelerated with greater efficiency. The work should be considered to be related to the earlier work reported in [9] [10].
ANN as a unitary classifier though generates a number of classification boundaries, but fail to reach optimality of decision making as per theoretical considerations for which Multiple Classifier System (MCS)s are preferred. These help to enhance the ability of the system to make decisions adjusting to finer variations. Such capability makes it relevant for the study of EAS because of the volume of correlated data involved. This work considers the use of several data dimension reduction techniques for use with a MCS constituted using Multi Layer Perceptron (MLP), Recurrent Neural Network (RNN) and Probabilistic Neural Network (PNN) for application in high energy shower analysis with special emphasis on prediction of EAS primary energy and core positions. The highly correlated density values provided by detectors distributed in a circular arrangement of 100 meters are subjected to dimensionality reduction using PCA, ICA and SOM approximation before applying to MCS for primary energy prediction and location of core position of certain EAS. The EAS events are assumed to be taking place inside the arc with the detectors recording the phenomena from all the locations within the arrangement. The requirement is to train the MCS with reduced dimension density sets from the detectors to enable them predict primary energies and core locations and produce a comparative performance measure as demonstrated by the two approaches. A few works related to the use of MCS for data analysis, pattern recognition and clustering is included here and cted between [11] and [16].

II. System Model

The system consists of a conceptual arrangement of detectors in a circle of radius 100 meters with density values of the shower events considered in groups taken from each of the four quadrants. The core positions are considered to be placed inside the 50 meter radius. The experimental set - up consists of a group of data mapping or dimensionality reducing blocks constituted by Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Self Organizing Map (SOM) approximators and three ANNs blocks. These ANN blocks are Multi Layer Perceptron (MLP), Recurrent Neural Network (RNN) and Probabilistic Neural Network (PNN) structures. The MLP, RNN and PNN forms the MCS system. The density of the detector size is taken to be 100 per quadrant to obtain more detailed description for the EAS but since the data shall be correlated PCA, ICA and SOM approximator blocks are used to map the most relevant portion for use with the MCS system. The complete system is depicted in Figure 1.
Let Dlr_i and Cr_i be the sets of density values and core positions respectively captured from a detector array placed in the circular arrangement. During the shower event a core may be placed within the 50 meter arc during which the sensors placed inside and outside the 50 meter circle act as detectors. The data captured from the detectors are density values of the showers and are applied to PCA, ICA and SOM approximator blocks. These reduce the density values to one fourth of the applied size which is equally contributed by each of the four quadrants of the circular arrangement.
The output of the SOMs are applied to the three ANNs blocks which are MLP, RNN and PNN structures. The output of these three blocks can be given as

$$y_{31} = y_{11} \times [ANN_1]|_{y_{21}} \tag{1}$$

$$y_{32} = y_{11} \times [ANN_2]|_{y_{21}} \tag{2}$$

$$y_{33} = y_{11} \times [ANN_3]|_{y_{21}} \tag{3}$$

where ANN_1 is and MLP, ANN_2 is a RNN and ANN_3 is a PNN and are trained as per the considerations mentioned in [10] [18].
The final segment of the system is a SOM block used as an optimizer of the outputs generated by the three ANNs. At a given instant the SOM retains the best output among the three ANNs. The selection is made by resorting to "Winner Takes All" approach of training and an Euclidean distance based cost function [18]. The optimization rule can be expressed as

$$Y_{out} = Best \;\; of\{y_{31}, y_{32}, y_{33}\} \tag{4}$$

III. Principal Component Analysis Approach of Dimensionality Reduction

Principal Components Analysis (PCA), is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. PCA is a powerful tool for analyzing data. Another advantage of PCA is that once these patterns are found, the data can be compressed, i.e. by reducing the number of dimensions, without much loss of information [19].

A. Definition of PCA:

A principal component can be defined as a linear combination of optimally - weighted observed variables. PCA is a classical statistical method and is a linear transform widely used in data analysis and compression. It is based on the statistical representation of a random variable [20].
Suppose we have a random vector population x, where

$$X = (x_1, ...x_n)^T \tag{5}$$

and the mean of that population is denoted by

$$\mu_x = E\{x\} \tag{6}$$

The covariance matrix of the same data set is

$$C_x = E\{(x - \mu_x)(x - \mu_x)^T\} \tag{7}$$

The components of C_x, denoted by c_{ij}, represent the covariances between the random variable components x_i and x_j. The component c_{ii} is the variance of the component x_i. The variance of a component indicates the spread of the component values around its mean value. If two components x_i and x_j of the data are uncorrelated, their covariance is zero (c_{ij}=c_{ji}=0).

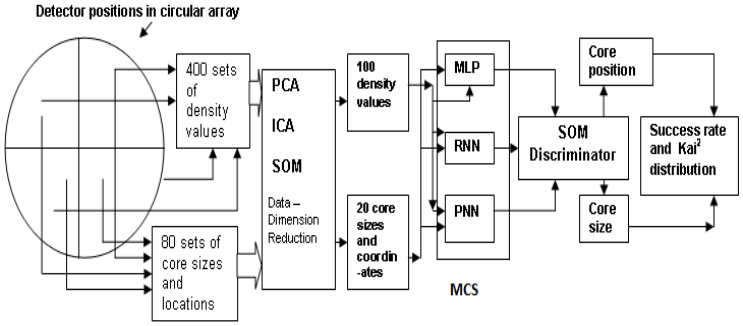

Fig. 1. System Model showing PCA, ICA and SOM approximators as part of MCS for EAS primary energy prediction and core location

The covariance matrix is, by definition, always symmetric [20]. From a sample of vectors x_1,.....,x_M , we can calculate the sample mean and the sample covariance matrix as the estimates of the mean and the covariance matrix [20].
From a symmetric matrix such as the covariance matrix, we can calculate an orthogonal basis by finding its eigenvalues and eigenvectors. The eigenvectors e_i and the corresponding eigenvalues λ_i are the solutions of the expression

$$C_x e_i = \lambda_i e_i, i = 1, ...n \quad (8)$$

For simplicity we assume that the λ_i are distinct. These values can be found, for example, by finding the solutions of the characteristic equation

$$|C_x - \lambda I| = 0 \quad (9)$$

where the I is the identity matrix having the same order than C_X and the |.| denotes the determinant of the matrix. If the data vector has *n* components, the characteristic equation becomes of order *n*. This is easy to solve only if *n* is small. Solving eigenvalues and corresponding eigenvectors is a non-trivial task, and many methods exist. One way to solve the eigenvalue problem is to use a neural solution to the problem. The data is fed as the input, and the network converges to the wanted solution [20]. By ordering the eigenvectors in the order of descending eigenvalues (largest first), one can create an ordered orthogonal basis with the first eigenvector having the direction of largest variance of the data. In this way, we can find directions in which the data set has the most significant amounts of energy [20].

B. Dimension Reduction using PCA

Principal component analysis is a widely used tool for dimension reduction. Let $x_i \in IR^d$, where i = 1,..., n, be the training patterns. The principal components are a set of $q < d$ orthonormal vectors and span a subspace in the major directions into which the patterns extend as in Figure 3 [21].

Let us assume that the patterns are centered around the origin (without loss of generality). Let y be the projection onto a subspace,

$$y = W^T x \quad (10)$$

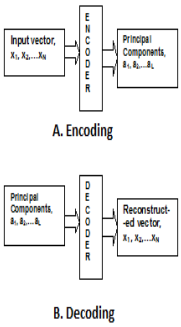

Fig. 2. Encoding-decoding process using PCA [18]

W is a dq matrix that contains the principal components as columns. The vector y is a dimension-reduced representation of x. Let $\hat{x}$ be the reconstruction of x given only the vector y,

$$\hat{x} = Wy \quad (11)$$

The goal of PCA is to set the subspace such that the mean reconstruction error E_{rec} is minimized'

$$E_{rec} = \frac{1}{n}\sum_{i=1}^{n} \|x_i - \hat{x}_i\|^2 \quad (12)$$

This goal is equivalent to finding the q major directions of maximal variance within the set of patterns x_i[21]. Moreover, it is equivalent to the principal components being the first q eigenvectors W_1 of the covariance matrix C of the pattern set,

$$C = \frac{1}{n}\sum_{i=1}^{n} x_i - x_i^T \quad (13)$$

The corresponding eigenvalue equation is

$$cw_1 = \lambda_l W_1 \quad (14)$$

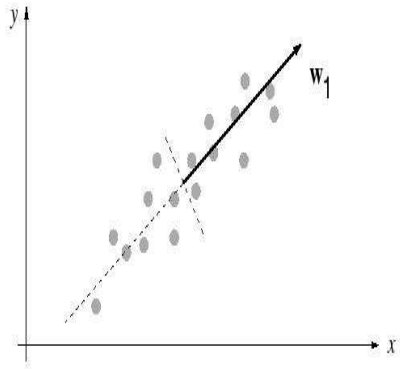

Fig. 3. The principal component points into the direction of maximum variance. The gray dots are the training patterns. The intersection of the dashed lines is the center of the pattern distribution[21]

The eigenvalue λ_l is the variance of the distribution x_i in the direction of W_1[21].

C. Generalized Hebbian Algorithm (GHA) Based PCA:

There is a close relationship between SOMs and PCA such that the Hebbian learning can be used to formulate a PCA type encoder. Let *m* be the number of inputs and *l* be the number of outputs such that *m¿l*. Let *[w]* be a matrix of random values representing the synaptic weights connecting inputs to the output layer. Thus for the set of inputs $x_i(n)$, the output is given as

$$y_j(n) = \sum_{i=1}^{m} w_{ji}(n)x_i(n) \quad (15)$$

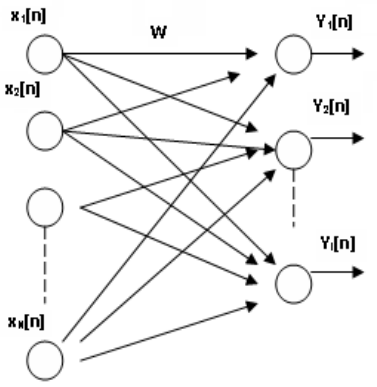

Fig. 4. GHA based PCA [18]

The weight matrix *[w]* is updated following a Generalized Hebbian Learning Algorithm expressed as

$$\Delta w_{ji}(n) = \eta[y_j(n)x_i(n) - y_j(n)\sum_{k=1}^{j} w_{ki}(n)y_k(n)] \quad (16)$$

where $\Delta w_{ji}(n)$ is the change in the synaptic weight matrix at time n and η is the learning rate parameter.

D. Adaptive Principal Component Extraction (APEX) Algorithm:

The Adaptive Principal Component Extraction (APEX) Algorithm is an extension of the GHA based PCA. It uses both feedforward and feedbackward connections. The algorithm calculate the first $j-^{th}$ principal components from the given (j-1)-components.

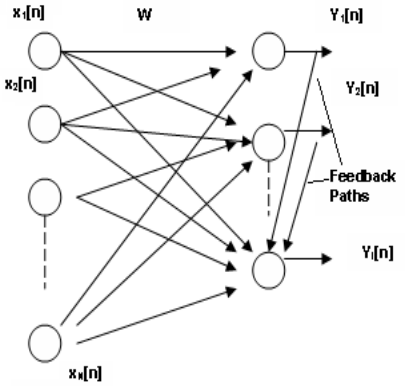

Fig. 5. APEX algorithm based PCA [18]

Let $\mathbf{w}_j = [w_{j1}(n), w_{j2}(n)...w_{jm}(n)]^T$ be the feedforward weight vector and $\mathbf{a}_j(n) = [a_{j1}(n), a_{j2}(n)...a_{jj-1}(n)]^T$ be the feedbackward weight vector. The feedforward part works with a Hebbian Learning Rule while the feedbackward part uses an anti-Hebbian updation [18]. The output $y_j(n)$ of the neuron j is given by

$$y_j(n) = w_j^T(n)x(n) + a_j^T(n)y_{j-1}(n) \quad (17)$$

The update equations for $\mathbf{w}_j$ and $\mathbf{a}_j(n)$ are expressed as

$$y_j(n) = w_j^T(n)x(n) + a_j^T(n)y_{j-1}(n) \quad (18)$$

The update equations for $\mathbf{w}_j$ and $\mathbf{a}_j(n)$ are expressed as

$$w_j(n+1) = w_j(n) + \eta[y_j(n)x(n) - y_j^2(n)w_j(n)] \quad (19)$$

and

$$a_j(n+1) = a_j(n) - \eta[y_j(n)y_{j-1}(n) + y_j^2(n)a_j(n)] \quad (20)$$

TABLE I
COMPUTATIONAL COMPLEXITY OF GHA AND APEX ALGORITHMS

Sl Num	Algorithm	Computational Complexity
1	GHA-PCA	$0.5(m^2+m)(p+1)+2pm$
2	APEX-PCA	$2pm+1.5m^2+2.5m$ m being the principal component, p being uncorrelated zero mean random process

E. Dimensionality Reduction using Kernal PCA:

Kernel principal component analysis (kernel PCA) is an extension of principal component analysis (PCA) using techniques of kernel methods. Using a kernel, the originally linear operations of PCA are done in a reproducing kernel Hilbert space with a non-linear mapping [22].

Different from the mixture models, kernel PCA just works with a single PCA. It is an extension of PCA to non-linear distributions. Instead of directly doing a PCA, the n data points x_i are mapped into a higher-dimensional (possibly infinite-dimensional) feature space, [21]

$$x_i \longrightarrow \varphi(x_i) \tag{21}$$

In the feature space, principal components are extracted. That is, the following equation needs to be solved where it is assumed that (x_i) has zero mean

$$\lambda W = CW \tag{22}$$

with the covariance matrix

$$C = \frac{1}{n}\sum_{j=1}^{n}\varphi(x_j)\varphi(x_j)^T \tag{23}$$

From the definition of C follows that C_W is a linear combination of the vectors $\varphi(x_i)$. Thus W, must lie in the span of $\varphi(x_1),...,\varphi(x_n)$ [21]. Hence, we can write

$$W = \sum_{i=1}^{n}\alpha_i\varphi(x_i) \tag{24}$$

Combining Eq 22 and Eq 24 gives

$$\lambda\sum_{i=1}^{n}\alpha_i\varphi(x_i) = \frac{1}{n}\sum_{i,j=1}^{n}\varphi(x_1)\alpha_i\varphi(x_j)(\varphi(x_j)^T\varphi(x_i)) \tag{25}$$

which is equivalent to the set of n equations

$$\lambda\sum_{i=1}^{n}\alpha_i(\varphi(x_i)^T\varphi(x_1)) = \frac{1}{n}\mathbf{XY} \tag{26}$$

where

$$\mathbf{X} = \sum_{i,j=1}^{n}(\varphi(x_j)^T\varphi(x_1)) \qquad \forall l \tag{27}$$

and

$$\mathbf{Y} = \sum_{i,j=1}^{n}(\varphi(x_j)^T\varphi(x_i)) \qquad \forall l \tag{28}$$

Using the kernel matrix, Eq 26 can be written as

$$n\lambda K\alpha = k^2\alpha \tag{29}$$

Thus, the vector α for each principal component can be obtained by extracting the eigenvectors of K. For further processing, the principal component W needs to be normalized to have unit length [21]. This can be also established by working solely with the kernel,

$$\|W\|^2 = (\sum_{i=1}^{n}\alpha_i\varphi(x_i)^T)(\sum_{j=1}^{n}\alpha_j\varphi(x_j)^T) = \alpha^T K\alpha = n^\lambda\alpha^T\alpha \tag{30}$$

which results in a normalization rule for α. To apply kernel PCA, a data point's features (the projections on the principal components) need to be extracted, and the formalism needs to be adjusted to distributions that do not have zero mean in feature space [21].

PCA itself is a powerful technique for extracting structure from possibly high-dimensional data sets. But it is not effective for data with non-linear structure. In kernel PCA, the input data with nonlinear structure is transformed into a higher-dimensional feature space with linear structure, and then linear PCA is performed in the high-dimensional space [23].

IV. INDEPENDENT COMPONENT ANALYSIS (ICA) APPROACH OF DIMENSIONALITY REDUCTION

ICA is a method in which the goal is to find a linear representation of non-Gaussian data so that the components are statistically independent, or as independent as possible. Such a representation seems to capture the essential structure of the data in many applications, including feature extraction and signal separation [25].

A. Definition of ICA:

To rigorously define ICA (Jutten and Hrault, 1991; Comon, 1994), one can use a statistical "latent variables" model. Let us assume that we observe n linear mixtures $x_1,...,x_n$ of n independent components

$$x_j = a_{j1}s_1 + a_{j2}s_2 + ... + a_{jn}s_n, \qquad for\ all\ j \tag{31}$$

The time index t has not been included. In ICA model, each mixture x_j as well as each independent component s_k is assumed to be a random variable, instead of a proper time signal. The observed values $x_j(t)$, are then a sample of this random variable with zero mean which makes the model zero-mean [25].

It is convenient to use vector-matrix notation instead of the sums like in the previous equation. Let us denote by x the random vector whose elements are the mixtures $x_1,...,x_n$, and likewise by s the random vector with elements $s_1,...,s_n$. Let us denote by **A** the matrix with elements a_{ij}. All vectors are understood as column vectors; thus x^T, or the transpose of x, is a row vector. Using this vector-matrix notation, the above mixing model is written as

$$x = As \tag{32}$$

Sometimes we need the columns of matrix **A**; denoting them by a_j the model can also be written as

$$x = \sum_{i=1}^{n}a_i s_i \tag{33}$$

The statistical model in Eqs 33 is called ICA model. The ICA model is a generative model, which means that it describes how the observed data are generated by a process of mixing the components s_i. The independent components are latent variables, meaning that they cannot be directly observed. Also the mixing matrix is assumed to be unknown. All we observe is the random vector x, and we must estimate both **A** and s using it. This must be done under as general assumptions as possible.ICA is very closely related to the method called blind source separation (BSS) or blind signal separation [25].

B. Dimensionality Reduction using ICA

Speech signals are composed of independent higher order statistical characteristics. ICA is often used in speech processing to extract the most important signal components [26]. Here an ICA network is trained to obtain independent components **u** from speech segment **x**, and the trained weight matrix **W** extract basis function coefficients **u** from **x**. ICA assume the observation **x** is the linear mixture of the independent components **u**. If **A** denote the inverse matrix of **W** then the columns of **A** represent basis feature vectors of observation **x** [26].

$$u = W.X_i, \qquad X = A.u \tag{34}$$

To extract basis functions one has to train mixing matrix **A** or unmixing matrix **W**, and we trained the mixing matrix **W**.
The learning rule is based on maximization of joint entropy **H(y)**, and is represented as

$$\Delta W \alpha \ \frac{\delta I(y,x)}{\delta W} = \frac{\delta H(y)}{\delta W} \tag{35}$$

$$\Delta W \alpha \ [W^T]^{-1} + \frac{\frac{\delta p(u)}{\delta u}}{p(u)} x^T \tag{36}$$

where **p(u)** denotes the approximation of the speech signal component probability density function,

$$p(u_i) = \frac{\delta y_i}{\delta u_i} = \frac{\delta g_i}{\delta u_i} \tag{37}$$

Here, **g(u)** is a nonlinearity function, which approximates the cumulative distribution function of the source signal **u** [26].
Natural gradient is also introduced to improve a converging speed. Particulary, this method does not require the inverse of matrix W, and provides the following rule:

$$\Delta W \alpha \ \frac{\delta H(y)}{\delta W} W^T W = [I - \varphi(u) u^T] W \tag{38}$$

where $\varphi(u)$ is related to the source probability density function and called as the score function [26].
Using the learning rule in Eqs 38, **W** is iteratively updated by gradient ascent manner until convergence. Let's denote **N** as the size of speech segments, which are randomly generated from training speech signals.
Figure 6 shows the basis vector training network. ICA network is composed of **N** inputs and **N** outputs, and **N** basis vectors are produced from **N** by **N** matrix **A (**$A = W^{-1}$**)** [26].
The above approach can be applied to extract the most important data sets from a host of samples provided by a grid of detectors arranged to monitor EAS events.

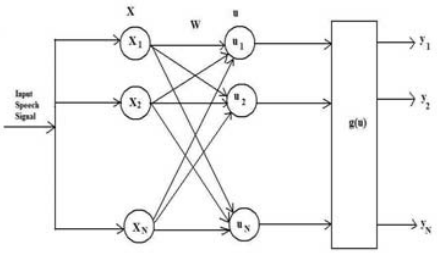

Fig. 6. ICA network for training the basis vector

C. Selection of Dominant Feature Vectors:

For a given set of correlated data, the most important ensemble can be selected from the **N** basis vectors. The ICA algorithm finds independent components corresponding to the dimensionality of the input, and may result in redundant components. To reduce this redundancy, several techniques have been proposed. The contribution of basis vectors to the input sample and the variability of the basis vector coefficients can be considered. The contribution means the power of the basis vector in the input samples and the L_2-norm ($\|a_i\|$, where a_i is the $i-^{th}$ column vector of **A**) can represent the relative importance of basis vectors. Therefore, from **N** basis vectors ordered in decreasing L_2-norm, **M** dominant and least correlated vectors can be selected. The variability denotes the variance of the basis vector coefficients, and this can represent the relative importance of basis vectors in extracting the most significant part of the input [26].

V. Experimental Details and Results

Experiments are carried out using density values taken from detector readings spread around a radius of 100 meters. The cores are assumed to be concentrated in an arc of 50 meter radius thus providing a set-up for derivation of density values using the NKG function [3]. These values are related to shower primary energy and the coordinates of the location where the event is assumed to have occurred.
The PCA, ICA and SOM approximator blocks at the input perform a process through which less than 25% of the samples supplied by the detectors are retained. These are data values are least correlated and can be considered to be provided by the detectors spread in a circular arc. In the true sense, the actual training data set comes from about 400 detectors which have high correlation. The highly correlated data can lead to inefficiency, hence the PCA, ICA and SOM approximator blocks are used which reduce the size of the input samples. The presence of the PCA, ICA and SOM approximator blocks helps in improving efficiency of the system as depicted by Table II. Experiments are carried out in several phases and the details are provided in [10].
The PCA based dimensionality reduction is carried using conventional PCA, GHA-PCA, APEX-PCA and Kernel-PCA.

TABLE II
AVERAGE IMPROVEMENT OF MCS PERFORMANCE DUE TO THE USE OF DATA APPROXIMATOR BLOCKS AT THE INPUT

Case	Epochs	Time in Sec.s	Success in Rate in %	Difference in time in sec.s	Difference in success rate in %
Without approximator blocks at Input	5000	87.2	92.0	-	-
	10000	107.3	93.0	-	-
	15000	132.5	93.0	-	-
With approximator blocks at Input	5000	72.0	93.0	14.9	1.0
	10000	93.0	95.0	14.3	2.0
	15000	111.8	95.0	20.7	2.0

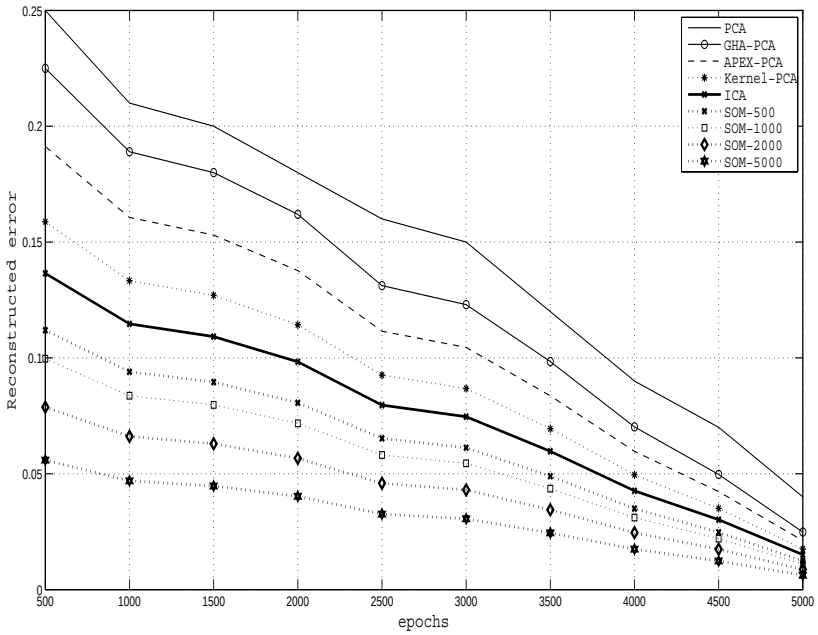

Fig. 7. Reconstruction error generated by PCA, ICA and SOM based data dimensionality reduction techniques

TABLE III
CONFIGURATION OF THE SOM APPROXIMATOR

Sl Num	Parameter	Specification
1	Input Size	400 x 100
2	Data Blocks	20
3	Multipaths	10
4	Output grid	1 x 100
5	Topology	*Gridtop* and *Hextop*
6	Distance function	*linkdist*
7	Ordering phase learning rate	0.5 to 0.9
8	Ordering phase steps	500 to 5000
9	Tuning phase learning rate	0.02 to 0.6
10	Epochs	100 to 5000 in steps of 100

Similarly, ICA based data approximation is carried out independently. The SOM approximators are used for data dimension reduction after 500, 1000, 2000 and 5000 sessions of training. The SOM approximator is configured as per the parameters given in Table III. Table III. A set of experiments are performed to determine the reconstruction error. The purpose is to ascertain the most suitable data dimensionality reduction technique for the MCS and Committee Machine based EAS analysis. The plot depicted in Figure 7 shows the reconstructed error when each of the PCA and ICA methods are implemented using respective ANN architectures. The Figure 7 also shows the reconstruction error generated by the SOM after four different training sessions. The SOM based data dimensionality reduction methods turn out to be the most suitable ones for the present application.

VI. CONCLUSION

This work describes certain data dimension reduction techniques like Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Self Organizing Map (SOM) approximators for application with an MCS formed using Multi Layer Perceptron (MLP), Recurrent Neural Network (RNN) and Probabilistic Neural Network (PNN). The data inputs are obtained from an array of detectors placed in a circular arrangement resembling a practical detector grid which have a higher dimension and greater correlation among themselves. PCA in four forms namely PCA, GHA-PCA, APEX-PCA and Kernel-PCA are considered. The coding decoding of

each of these four techniques are carried out using specific ANN architectures. Similarly, ICA based data dimensionality reduction is implemented. The work also investigates the effectiveness of the SOM as an approximator at the end of different training sessions. The SOM based approximation provides the best results in terms of reconstruction error though PCA based techniques showed better computational complexity than ICA. The techniques of data dimensionality reduction discussed here provides an insight regarding how the best set of density values can be considered for EAS analysis using a MCS. The result is a system with enhanced features of data dimensionality reduction suitable for applications for primary energy prediction and core location detection proving to be effective for EAS analysis.

REFERENCES

[1] R. A. Mewaldt: "Cosmic Rays", Macmillan Encyclopedia of Physics- available at http://www.srl.caltech.edu/personnel/dick/cos-encyc.html, 1996.
[2] R. Engel: "Theory and Phenomenology of Extensive Air Showers", Forschungszentrum Karlsruhe, Germany- available at http://moriond.in2p3.fr/J05/trans/sunday/engel1.pdf, 2005.
[3] D. Hanna, "Application of Neural Nets to Extensive Air-Showers", Proceedings of the 22nd International Cosmic Ray Conference, IUPAP, Volume 4, p: 500, 1991.
[4] J. C. Perrett and J T P M van Stekelenborg, "The applications of neural networks in the core location analysis of extensive air showers", Bartol Research Institute, University of Delaware, 217 Sharp Laboratory, Newark, DE 19716, USA J. Phys. G: Nucl. Pert. Phys. 17, pp:1291-1302, 1991.
[5] H. J. Mayer, "A neural network algorithm for core location analysis at large extended air shower arrays", Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 317, Issues 1-2, pp. 339-345, June, 1992.
[6] T. Wibig, "The Artificial Neural Networks as a tool for analysis of the individual Extensive Air Showers data", Preprint submitted to Elsevier Preprint (1 February 2008), available at http://arxiv.org/abs/hep-ph/9608227v1.
[7] P. Sommers: "Extensive Air Showers and Measurement Techniques", Physics - Astrophysics, C. R. Acad. Sci. Paris, t. 4, Serie IV, pp. 1 - 12, 2004.
[8] A. Chilingarian, G. Gharagyozyan, S. Ghazaryan, G. Hovsepyan, E. Mamidjanyan, L. Melkumyan, V. Romakhin, A. Vardanyan and S. Sokhoyan: "Study of extensive air showers and primary energy spectra by MAKET-ANI detector on mountain Aragats", Elsavier Journal of Astroparticle Physics, vol. 28, pp. 5871, 2007.
[9] G. Devi, K. K. Sarma, P. Datta and A. K. Mahanta: "Prediction of Location of High Energy Shower Cores using Artificial Neural Networks", International Journal of Engineering and Applied Sciences Volume 7, No.1, pp.-20 - 26, 2011.
[10] G. Devi, K. K. Sarma, P. Datta and A. K. Mahanta: "ANN based Multi Classifier System for Prediction of High Energy Shower Primary Energy and Core Location", International Journal of Information and Mathematical Sciences, Vol 6, No. 3, pp. 176- 185, 2010.
[11] L. Chen and M. S. Kamel: "Design of Multiple Classifier Systems for Time Series Data ", Spinger Lecture Notes in Computer Science, 2005, vol. 3541, pp. 216-225, 2005.
[12] M. T. El-Melegy and S. M. Ahmed: "Neural Networks in Multiple Classifier Systems for Remote-Sensing Image Classification", Spinger Studies in Fuzziness and Soft Computing, vol. 210, pp. 65-94, 2007.
[13] Z. Suraj, N. E. Gayar, P. Delimata: "A Rough Set Approach to Multiple Classifier Systems", Fundamenta Informaticae, vol. 72, no. 1-3, pp. 393-406, 2006.
[14] F. Roli: "Mini Tutorial on Multiple Classifier Systems", University of Cagliari, Dept. of Electrical and Electronics Eng., Italy, School on the Analysis of Patterns, 2009.
[15] M. Sharma and R. Mammone: "Speech recognition using sub-word neural tree network models and multiple classifier fusion", CAIP Center, Rutgers University, Piscataway, NJ, USA, 1995.
[16] M. Sharma and R. Mammone: "Artificial neural network fusion: Application to Arabic words recognition", Proceedings of ESANN - 2005- European Symposium on Artificial Neural Networks Bruges, (pp. 151 - 156), Belgium, 27-29 April, 2005.
[17] B. Yegnanarayana, *Artificial Neural Networks*, 1^{st} Ed., PHI, New Delhi, 2003.
[18] S. Haykin, *Neural Networks A Comprehensive Foundation*, 2^{nd} Ed., Pearson Education, New Delhi, 2003.
[19] L. I. Smith: A tutorial on Principal Components Analysis: February 26, 2002.
[20] S. Fiori: "An Experimental Comparison of Three PCA Neural Networks", Neural Processing Letters, vol. 11, pp. 209-218, 2000.
[21] H. Hoffmann: "Unsupervised Learning of Visuomotor Associations: Dissertation", Universitt Bielefeld, Technische Fakultt, Logos Verlag Berlin, ISBN: 3-8325-0858-9, Germany, 2005.
[22] Kernel principal component analysis: Wikipedia, the free encyclopedia $/en.wikipedia.org/wiki/Kernel_principal_component_analysis$
[23] T. Takiguchi, Y. Ariki: "PCA-Based Speech Enhancement for Distorted Speech Recognition", Journal of Multimedia, vol. 2, no. 5, 2007.
[24] I. Rosca, L. State and C. L. Cocianu: "Investigations on the Potential of PCA Based Neural Implementation Attempts in Solving Specific Tasks in Image Processing", 9th WSEAS Int. Conf. on Mathematics and Computer in Business and Economics (MCBE '08), Bucharest, Romania, pp. 116-124, June 24-26, 2008.
[25] A. Hyvrinen and E. Oja: "Independent Component Analysis- Algorithms and Applications: Neural Networks", Research Centre, Helsinki University of Technology, Finland, 13 (4-5), pp. 411-430, 2000.
[26] J. H. Lee, H. Y. Jung, T. W. Lee, S. Y. Lee: "Speech Feature Extraction Using Independent Component Analysis", In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00), Istanbul, Turkey, vol.3, pp. 1631 - 1634, 2000

G itanjali Devi is presently with the Department of Physics, Lalit Chandra Bharali College, Guwahati-781011, Assam, India. Her fields of interest are ANN and its application for EAS and high energy particle analysis.

K andarpa Kumar Sarma is presently with the Department of Electronics and Communication Technology, Gauhati University, Guwahati - 781014, Assam, India. His areas of interest include Applications of ANNs and Neuro-Computing, Document Image Analysis, 3-G Mobile Communication and Smart Antenna.

P rof. (Mrs) Pranayee Datta is presently the Head, Deptt. of Electronics and Communication Technology and Dean, Faculty of Technology, Gauhati University, Guwahati - 781014, Assam, India. Her areas of interest include Radio Astronomy, High Energy Cosmic Rays, Nano Science and Technology, Antenna Design.

P rof. (Mrs) Anjana Kakoti Mahanta is presently the Head, Deptt. of Computer Science, Gauhati University, Guwahati - 781014, Assam, India. Her areas of interest include Algorithms and Data Structure, Data Mining and Warehousing.