STEP 7: Step 2 is performed and calculates the accuracy (correctly classified genes) till the end of total number of iterations.
STEP 8: Based on the final iterations result, we identify the minority and majority classes
STEP 9: Perform the under sampling, which reduces the size of the data of majority class equal to size of the data of minority class or Perform the over sampling, which replicates synthetic data to the minority class equal to size of the data of majority class or Perform the smote sampling, which adds new synthetic data to the minority class equal to size of the data of majority class.

Despite, the evolutionary sampling method generates the population that are fittest for classification. After balancing the dataset through sampling the outcome of the technique further undergoes optimality.

### 3.3 Particle Swarm Optimization Technique

Particle swarm optimization (PSO) is a population based stochastic optimization technique (M. Saberi et al, 2009) proposed by Dr. Eberhart and Dr. Kennedy in 1995. PSO shares many similarities with evolutionary Genetic Algorithms (GA). The system is initialized with a population of random solutions and searches for optimal solution by updating generations. However, unlike Genetic algorithm, PSO does not constitute genetic operators such as crossover and mutation. In PSO, the solutions, called particles, flutter through the problem space by following the current optimum particles.

In this work, each particle of PSO keeps track of its coordinates in the problem space. The co-ordinates are the particles which are associated with the best solution called fitness. (The fitness value is also stored.) This value is called pbest. Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained by any particle related to the neighbors of the particle. This location is called lbest. Further, when a particle takes all the population of its topological neighbors, the best value is termed as global best called gbest. Formulating the best position and velocity is the key aspect in this work.

The particle swarm optimization in each iteration, changes the velocity of (accelerating) each particle toward its pbest and lbest locations. It is demonstrated that PSO gets better results in a cheaper, faster way when compared with other methods. The algorithm for PSO proceeds as follows.

**Notation 1:** Let $f^{\circ} = D^n \rightarrow D$ is the condition that has to be minimized from$D^n$ to D

**Parameters:** Velocity $v_i$ and the position $p_i$ , local best$l_b$, Particle best $p_b$ and global best $g_b$,$(b_l, b_u)$ are the upper and lower boundaries of the search space and are referred as parameters of PSO.

STEP 1: For each particle i=1,2………S initialize the particles position

$$p_i \sim \cup\ (b_l, b_u)$$

STEP 2: Initialize the particle best position to its initial position $p_b \leftarrow p_i$.
STEP 3: If $(f(p_b) < f(g_b))$ the swarm's best known position is updated to $g \leftarrow p_b$
STEP 4: Initialize the particle velocity

$$v_i \sim \cup\ (-|b_u - b_l|, |b_u - b_l|)$$

STEP 5: For each particle i= 1,2,3……………S then do
For each dimension $D_d = 1,2 \ldots \ldots \ldots \ldots \ldots n$ do
STEP 6: Extract random numbers such that $rp_b rg_b \sim \cup$ (0,1)
STEP 7: Update the particle's velocity

$$v_{id} \leftarrow m(gf_{ij})^A v_{id} + gf_p r_p (p_{bd} - p_{id}) + gf_g r_g (g_{bd} - p_{id})$$

STEP 8: Update the particle's position

$$p_i \leftarrow (p_i + v_i)$$

STEP 9: if $(f(p_b) < f(g_b))$ update the swarm's best known positioning$g_b \leftarrow p_b$.
STEP 10: Now $g_b$ holds the best found solution.

Thus PSO selects the optimal features for classification. It is a search heuristic method that selects the optimal feature for classification.

### 3.4 SVM Classification

The outcome of the balanced dataset is classified dataset using SVM classification. After classification, the data that lies on the above margin are the features that are highly satisfied. Hence these features are selected for prediction analysis. The SVM classification is used to relate the data items and to select the appropriate features for prediction.

## 4. Experimental Results

The experiment is carried on three datasets of micro array they are Lung cancer, Lymphoma and colon cancer. The dimensions of the dataset are 4071x96, 2026x96, 2026x96 instances and features. The performance of the proposed work is evaluated using the accuracy parameter on the sampling techniques.

**Table 1.1:** Result

| Accuracy % | | | |
|---|---|---|---|
| Technique/ Dataset | Over Sampling | Under Sampling | SMOTE Sampling |
| Lymphoma | 93.7 | 84.6 | 85.6 |
| Lung cancer | 94.3 | 88.6 | 86.6 |
| Colon | 92.4 | 82.5 | 85.9 |