

# A novel approach for detecting sensor-based semiconductor fault yield classification using convolutional neural networks

Mohammed Altaf Ahmed, Suleman Alnatheer

Department of Computer Engineering, College of Computer Engineering and Sciences,  
Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia

## Article Info

### Article history:

Received Oct 24, 2023

Revised Dec 29, 2023

Accepted Jan 11, 2024

### Keywords:

Convolutional neural network

Greedy spectral selection

Inter band redundancy

Semiconductor fault

Yield classification

## ABSTRACT

In the proposed research, data from the semiconductor industry is considered for analysis. In this research, there is a requirement for significantly more space for storage, processing will take significantly more time, and there will be a significant amount of duplicate data. So, the utilization of dimensionality reduction strategies is required so as to lessen the number of spectral bands while maintaining the maximum amount of relevant information. Our contribution can be broken down into two parts: To begin, we suggest a filter-based technique that we call interband redundancy analysis (IBRA). This method is based on a collinearity analysis that is performed among a band and its neighbors. By performing the given research, redundant bands can be omitted, which in turn significantly brings down the search space. Next, we take the findings of the IBRA and use a wrapper-based technique known as greedy spectral selection (GSS) to choose bands on the basis of the information entropy values of those bands. We are later training a convolutional neural network to evaluate how well the present selection is working. We also propose an optimization algorithm for performance enhancement known as bacterial foraging optimization.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## Corresponding Author:

Mohammed Altaf Ahmed

Department of Computer Engineering, College of Computer Engineering and Sciences

Prince Sattam bin Abdulaziz University

Al-Kharj, Saudi Arabia

Email: m.altaf@psau.edu.sa

## 1. INTRODUCTION

Analyses are performed on the data obtained from the semiconductor manufacturing industry, which contains information that is both spatial and spectral in nature [1]. When identifying objects in a scene, it is necessary, in a number of such applications, to take into account the spatial information that is being carried in the image; nevertheless, the spectral information plays an important part in this procedure [2]. The capacity to maintain spatial resolution while simultaneously producing a multitude of spectral channels has led to the development of a wide variety of imaging systems for passive remote sensing. These systems are able to preserve spatial resolution. These systems range from multispectral imaging (MSIs) [3], [4] to hyperspectral imaging systems (HSIs) [5], [6]. HSIs can capture high spatial resolution while simultaneously recording a wide range of content in spectral. There is a possibility that HSI systems' added complexity and expense are not required for certain applications, particularly those with spectral properties that fluctuate gradually. The determination of which wavelengths in a data cube are the most important is a challenging problem brought on by the fact that HSI systems collect a large amount of spectral data. If you were able to successfully determine the prominent wavelengths contained inside a hyperspectral data cube, you would be eligible for a number of different advantages [7]–[9]. The processing and storage needs can be simplified

when applying a band selection method since only important spectral bands are saved and processed [10]. While applying techniques for feature extraction, each spectral band is saved, but the accuracy of classification can be improved [11]. This is true regardless of which technique multispectral imagers is utilised to identify the significant spectral bands. This has many different utilizations, with a range from precision agriculture [12] to dermatological [13].

In this article, we describe an interband redundancy analysis (IBRA) dimensionality reduction technique. Our technique simplifies spectrally data recorded from the semiconductor processing industry by removing redundant spectral bands. Each spectral band and its neighbours are subjected to a recursive collinearity analysis in IBRA. This analysis is undertaken in order to determine whether or not the bands overlap. This enables us to calculate an approximation of the quantity of spectral bands that must be moved far from a band for finding spectral bands that include data that is sufficiently distinct from the information contained in other spectral bands. After the distance metric has been computed for each spectral band, it is utilised to cluster into sets of spectral bands the spectrum that are similar to one another and that are adjacent to one another by looking at how the distance metric is distributed across the spectrum. Because of this, we are able to identify a more manageable number of distinct bands, each of which can act as the spectrum's corresponding cluster's centroid.

Along with the simplification of the hyperspectral data cubes, the proposed IBRA technique could be utilised as part of a new dual-step hybrid feature selection process. The initial step employs IBRA to identify of independent representative bands a reduced set; then, as shown in Figure 1, it employs the proposed wrapper-based technique known as greedy spectral selection (GSS) for selecting a user-defined quantity of spectral channels. GSS assigns a score to each candidate band based on the information entropy of that band in order to generate an initial selection of bands. This bands selection is then used for training a classifier, which finds out the efficacy of the supplied bands selection on the basis of the classification quality it produces. This evaluation is carried out in order to determine how effective the supplied bands selection is. The band that represents the highly important sign of multicollinearity is then excluded from the selection process. The operation is then repeated by regarding the band with the largest information entropy, and it is checked to see if the classification efficiency increases. As a consequence of this, the goal of GSS is to maximize the discrimination among classes while still maintaining spectral uniqueness [14]. The given band selection procedure is used to assist us in accomplishing our objective of developing high-accuracy, low-cost MSI systems that can be utilised for a variety of different purposes.

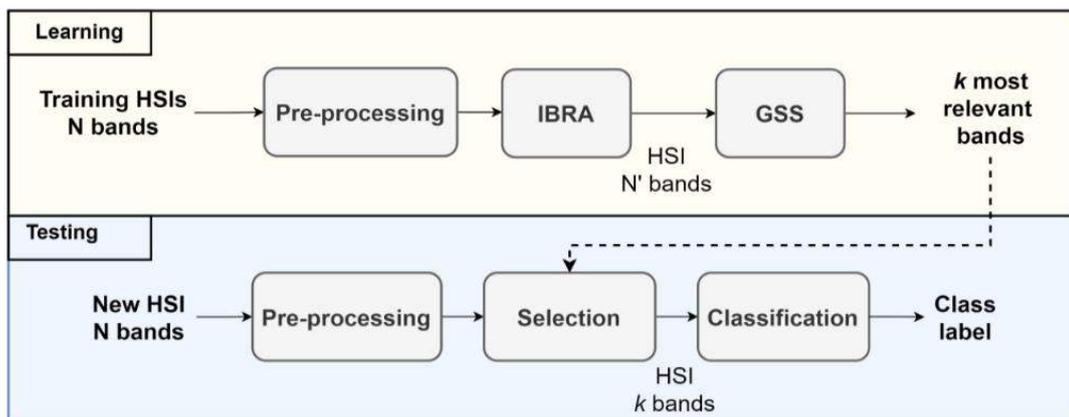


Figure 1. Flowchart utilising IBRA-GSS of the complete selection of band process

In addition, we investigate the viability of using IBRA in conjunction with similar dimensionality reduction approaches in addition to the many feature selection methods that are currently available. We will explain a method for the extraction of features, the first stage of which is to eliminate spectral redundancy by the use of IBRA. The second stage involves applying any dimensionality reduction technique wanted to the set of spectral bands that were chosen by the IBRA in order to build a new representation with the necessary feature channels or dimensions. As a consequence of this, the proposed framework possesses a structure that is analogous to the one depicted in the figure of flowchart, although it does not contain a GSS block but rather a dimensionality reduction technique. In the second step of this paper, we present experimental evaluations for this framework utilising conventional neural network technique. The given methods is used to predict the performance of the proposed research. In the context of spectral bands analysis of semiconductor

yield, the findings indicate that the suggested method for prediction of performance generates results that are superior to or on par with those produced by many alternative methods for the selection of features.

## 2. LITERATURE REVIEW

Explaining when working with data from the semiconductor manufacturing industry, it is often beneficial to apply methods for reducing dimensionality as a preprocessing step. This helps to avoid an unnecessarily high level of computational complexity and reduces the amount of redundant data. Methods for reducing dimension not only lower the present computational cost and lessen the requirement for processing the data for advanced hardware specifications, in addition they fight the "curse of dimensionality" [15], [16]. The "curse of dimensionality" refers to the phenomenon in which the efficiency of an HSI classifier declines as the number of features in the HSI data set rises, and the smaller training samples there are. Dimensionality reduction strategies lower the inherent computing cost of processing the data and lessen the necessity of using more complex hardware.

Methods for reducing dimension depend on either feature selection or feature extraction strategies; the erstwhile utilise nonlinear or linear transformations for the extraction of particular features from the initial data, whilst the next one picks a most beneficial subset of the data features inspite of transforming them in any way. Dimensionality reduction techniques can be broken down into three categories: feature extraction, feature selection, and dimension reduction [17]. Both unsupervised and supervised feature extraction approaches can be broken down further into their respective categories. Principal component analysis (PCA) and related variants, for instance, kernel PCA and folded-PCA, are a few of the widespread utilised unsupervised techniques [18] for reducing the dimensionality and removing spectral redundancy of the raw data. PCA and its variants are also known as principal component analysis. PCA and its variants are also among the most popular methods for reducing the number of dimensions.

Despite this, both theoretical explanations and experimental data indicating that PCA is useless for the extraction of hyperspectral features were presented [19]. In view of these, PCA shouldn't be employed as a preprocessing step for solving issues regarding small sample sizes. There is indication that class separation declines after PCA transformation, so using PCA to resolve issues involving small sample sizes is not recommended.

Independent component analysis (ICA) and related variants, such as random fourier feature-ICA and Kernel ICA, serve as the basis for a variety of additional types of unsupervised approaches [20]. Alternatively, supervised techniques such as partial least squares discriminant analysis (PLS-DA) [21] or Fisher's linear discriminant analysis (LDA) [22], [23] are utilised rather commonly for the aim of hyperspectral feature extraction. Newer methods of feature extraction to locate nonlinear embeddings use local manifold learning, which are then employed for projecting the data into lower-dimensional spaces. Hong *et al.* [24] presented, for instance, a dependable local manifold description which is incorporating a hierarchical neighbour selection. This representation's goal is to reduce the amount of multicollinearity that exists in the locally present manifold area. Additionally, the weights that are computed are created in such a way as to concurrently incorporate information pertaining to both the spectral and spatial domains. It is also possible to use methods of feature extraction with the intention of improving the quality of the data. Partitioned relief-F, introduced by Ren *et al.* [25], is a better way to reduce the negative impact of continuous bands on classification accuracy without losing any useful data. Nevertheless, techniques for selecting features select a subset of spectral bands without making any changes to the data or projecting it into a new basis. A greater analysis of the materials of its optical properties can also be attained through the identification of a limited quantity of essential spectral bands, which also provides information that is helpful in the development of low-cost multispectral imagers that are tailored to a particular application. In addition to this a method of testing memory by the hardware level using the inbuilt controller [26] is presented in the research.

## 3. METHOD

In this study, we introduce an IBRA dimensionality reduction strategy. This method eliminates duplicated spectral bands, which in turn helps to reduce the amount of data from the semiconductor industry regarding semiconductor yield. IBRA is on the basis of a recursive collinearity analysis among every spectral band and its neighbours. This analysis gives us an approximation of the quantity of bands we require to shift away from a band for the location of spectral bands that contain sufficient information that is not present in any other band. After it has been calculated for each spectral band, the next step is to identify how this distance metric is distributed across the spectrum. This allows the spectrum to be grouped into sets of spectral bands that are similar to one another and that are next to one another. Because of this, we are able to zero in

on a more manageable number of separate bands that act as the spectrum's centroid for the clusters to which they are connected. An IBRA filter-based technique is shown first. This technique is on the basis of a collinearity analysis that is performed between a band and its neighbours. Because of this technique, duplicated bands may be taken out of the equation, which considerably limits the search space. Isometric mapping (ISOMAP), laplacian eigenmaps (LE), locally-linear embedding (LLE), locality preserving projection (LPP), and many more non-linear dimensionality reduction methods based on manifold learning have all been suggested and used to HSI data. When applied to HSI data, these non-linear dimensionality reduction techniques likewise undergo similar shifts.

Next, a wrapper-based technique is utilized that is called GSS to pick bands according to the information entropy values of those bands. A compact convolutional neural network is trained next for the evaluation of the effectiveness of the present selection. We also provide a framework for feature extraction, which is made up of two basic parts: first, it uses IBRA to reduce the overall quantity of bands, and second, it can use GSS technique for acquiring the suitable number of feature channels. Both of these steps are outlined below. The neural network used is convolutional neural network (CNN) and for optimization of results, bacterial foraging optimization algorithm is used. Experimental findings for this framework are shown here, with a filter-based technique that is known as IBRA. Our experiments with GSS reveal that our feature selection framework outperforms or is on par with other feature selection approaches.

### 3.1. Interband redundancy analysis

IBRA is a type of analysis that chooses a selection of typical spectral bands from the initial data cube with the intention of lowering the interband redundancy. IBRA is a filter-based selection technique that verifies the presence of collinearity between pairs of bands by iteratively calculating the variance inflation factor (VIF) among pairs of bands. This is done so as to determine how correlated the bands are with one another, or to establish whether or not they are collinear.

We created an ordinary least squares (OLS) regression model to calculate the value of the VIF that exists between two bands,  $x_1$  and  $x_2$ . One of the bands serves as the independent variable ( $x_1$ ) in this model, and the other serves as the dependent variable ( $x_2$ ). After that, we calculated the VIF value by using the R-squared (coefficient of determination) value that the resultant model provided, which was  $R^2_{x_1, x_2}$ .

$$VIF(x_1, x_2) = 1/(1 - R^2_{x_1, x_2})$$

The values of VIF range from one to infinity. The likelihood of collinearity increases in proportion to the value of the VIF variable. When dual bands are collinear, it means that they both explain inside the dataset the similar amount of variance, which means that they are redundant. When we look for collinearity between two spectral bands, we will assume that there is collinearity if the VIF values are higher than a certain threshold called. Because the recommended values of in the scientific literature range between 5 and 10, we conduct experiments with a variety of values of [5, 12] to determine how the classification performance is impacted by this threshold and to select the value of that is optimal for the classification task in question.

### 3.2. Greedy spectral selection

The first step in the GSS process involves assigning a ranking to every candidate band  $x_c$ , where  $x_c \in S_c$ , based on certain criterion. Within the scope of this study, we determined for every spectral band an initial relevance score by employing the information entropy measure. The entropy of band  $x_c$ , also known as  $H(x_c)$ , expresses the typical amount of information that is contained inside  $x_c$  and can be expressed as follows:

$$H(x_c) = -\sum_{z \in \Omega_{x_c}} P(z) \log P(z)$$

where  $\Omega_{x_c}$  is the space which is encompassing every possible values which could be occurring in the spectral band  $x_c$  and  $P(z)$  is the probability mass function of random variable  $z$ .

### 3.3. Convolutional neural network

Conventional artificial neural networks (ANNs) and CNNs are analogous to one another in the respect that both are composed of neurons that are capable of enhancing their own performance through the learning process. Every neuron in an artificial neural network still must be given an input and carry out some kind of action (like a scalar product then a non-linear function), as this is the fundamental component of an extraordinarily wide range of different ANNs. Starting with the raw image vectors that are input and concluding with the class score that is computed, the complete network will continue to display a single

perceptive scoring function. This begins with the input of the raw image vectors and ends with the calculation of the class score (the weight). Loss functions that are connected to the classes will be included in the most superficial layer and all of the normal recommendations and procedures that have been developed for conventional artificial neural networks will still be relevant.

### 3.4. Bacterial foraging optimization

The unique nature-oriented method known as bacterial foraging optimisation (BFO) [1] was developed by Passino. Its primary purpose is to imitate the behaviours of *Escherichia coli* as it goes about the procedure of looking for nutrients. In the area of food acquisition, bacteria will often engage in four major activities: chemotaxis, reproduction, elimination, and dispersal. In addition to this, the location of a single bacterium could be considered in the search region as a workable solution when it comes to optimising the function of the system. Bacteria are able to fine-tune their places by themselves by swimming in predetermined step sizes and tumbling in random directions. This allows them to continually locate the best possible spot. The BFO algorithm has steadily been gaining in popularity, and it has been utilized in a wide number of practical domains up until this point, including training kernel extreme learning machine, feature selection, and facility layout, to name a few. The BFO algorithm's gradual rise to popularity can be attributed to its many benefits, including its high level of robustness and its good performance in local search.

### 3.5. Dataset used

The dataset used for the proposed research is data from a semi-conductor manufacturing process. Monitoring the signals and variables that are gathered from various sensors and/or process measurement sites is typically done in order to keep a sophisticated modern manufacturing process for semiconductors under close and constant observation. Nevertheless, not each of these signals contributes the same amount of useful information to a particular monitoring system. The signals that were measured include noise in addition to helpful information and information that is not relevant to the situation.

The data in the dataset has multivariate data set characteristics. The number of instances is 1,567 and it has real attribute characteristics. The number of attributes is 591. The rows represent signals and columns represent bands. Every signal has a number of bands [27]. The sample dataset is given in Figure 2. The shape of the dataset is (1,567, 560). Statistical details of the dataset is given in Figure 3. Sensor semiconductor different signal band transmission ratio (sample) is represented in Figure 4.

Time	0	1	2	3	4	5	6	7	8	...	581	582	583	584	585	586	587
0 2008-07-19 11:55:00	3030.93	2564.00	2187.7333	1411.1265	1.3602	100.0	97.6133	0.1242	1.5005	...	NaN	0.5005	0.0118	0.0035	2.3630	NaN	NaN
1 2008-07-19 12:32:00	3095.78	2465.14	2230.4222	1463.6606	0.8294	100.0	102.3433	0.1247	1.4966	...	208.2045	0.5019	0.0223	0.0055	4.4447	0.0096	0.0201
2 2008-07-19 13:17:00	2932.61	2559.94	2186.4111	1698.0172	1.5102	100.0	95.4878	0.1241	1.4436	...	82.8602	0.4958	0.0157	0.0039	3.1745	0.0584	0.0484
3 2008-07-19 14:43:00	2988.72	2479.90	2199.0333	909.7926	1.3204	100.0	104.2367	0.1217	1.4882	...	73.8432	0.4990	0.0103	0.0025	2.0544	0.0202	0.0149
4 2008-07-19 15:22:00	3032.24	2502.87	2233.3667	1326.5200	1.5334	100.0	100.3967	0.1235	1.5031	...	NaN	0.4800	0.4766	0.1045	99.3032	0.0202	0.0149

5 rows × 592 columns

Figure 2. Sample dataset

	count	mean	std	min	25%	50%	75%	max
0	1561.0	3014.452896	73.621787	2743.2400	2966.260000	3011.4900	3056.6500	3356.3500
1	1560.0	2495.850231	80.407705	2158.7500	2452.247500	2499.4050	2538.8225	2846.4400
2	1553.0	2200.547318	29.513152	2060.6600	2181.044400	2201.0667	2218.0555	2315.2667
3	1553.0	1396.376627	441.691640	0.0000	1081.875800	1285.2144	1591.2235	3715.0417
4	1553.0	4.197013	56.355540	0.6815	1.017700	1.3168	1.5257	1114.5366

Figure 3. Statistical details

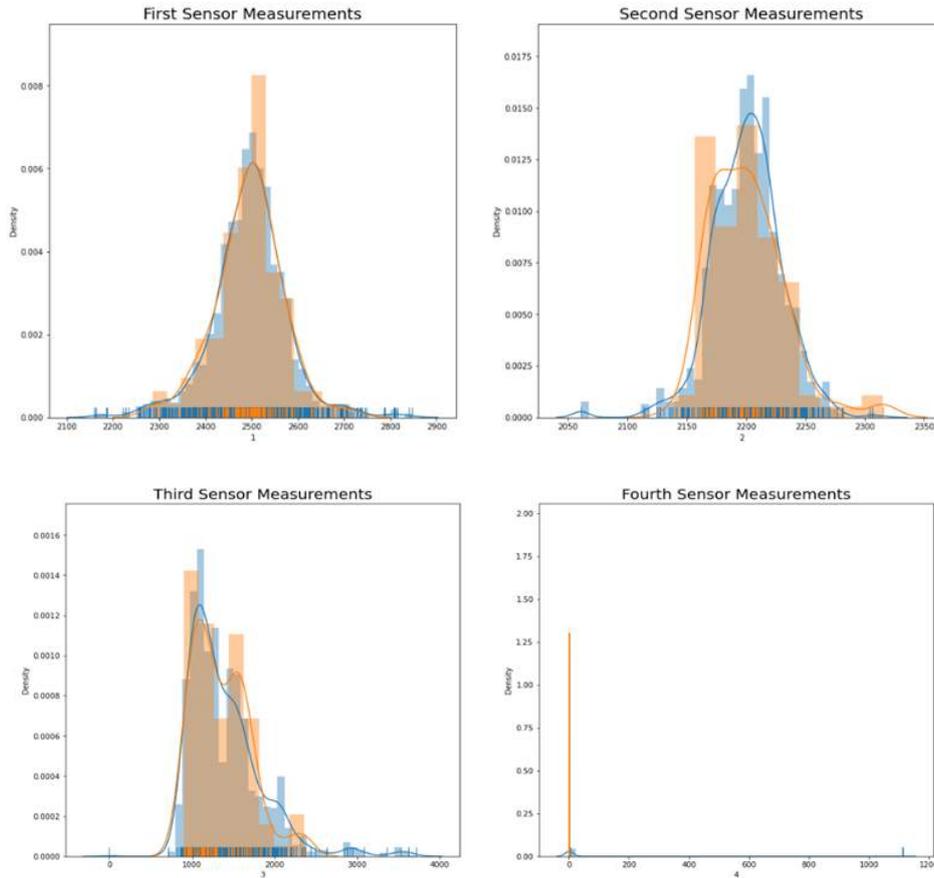


Figure 4. Sensor semiconductor different signal band transmission ratio (sample)

## 4. RESULTS AND DISCUSSION

### 4.1. Pre-processing

The first process in the preprocessing step is checking the null values. Null value check is (32,). Next check number of unique values on every signal bands. The input target type details are having a float64 as 590. Pass/failed yield for the sensor signal line and its target normalization is represented as (-1: 1463, 1: 104). -1 is used to represent pass yield which is 1,463 and 1 is used to represent fail yield which is 104. Large class imbalance can be seen. Since there were large number of missing values, some features have to removed manually. Graphical representation of distribution based on target data using pie chart and histogram is represented in Figure 5.

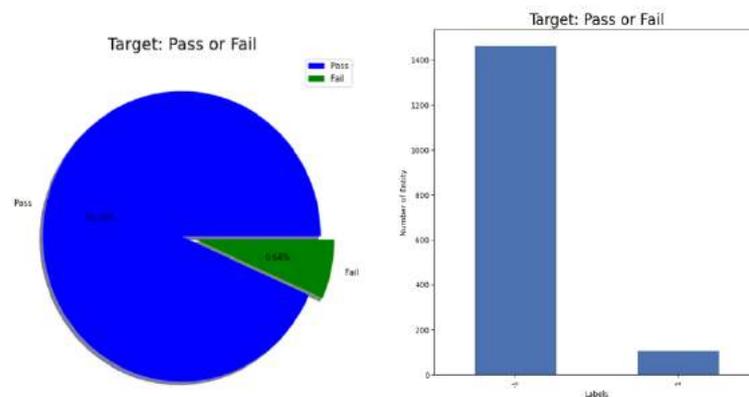


Figure 5. Distribution based on target data

Next, check non-unique faulty values and it can be seen that from the 560 remaining features, 116 of them have only 1 unique value which indicates a faulty signal across all batches and therefore these are dropped from the dataset. Shape after dropping null and faulty values is (1,567, 444). The missing signal points are checked next, and it is found that there are 8008 missing signal points. Signal missing label matrix graph (printing all our signals and checking where the signals are missing) is represented in Figure 6. There are a number of columns that contain the missing values. These columns are later dropped. Shape after dropping columns containing missing signal bands is (1,567, 423).

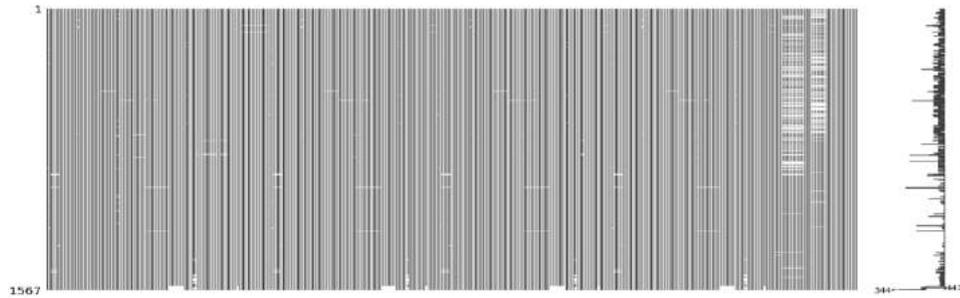


Figure 6. Signal missing label matrix graph

Since the above graphs are still showing some missing points, deep checking missing signal point using KNNImputer algorithm is done. Imputation is the process of employing k-nearest neighbors (KNN) to complete missing value information. Using KNNImputer algorithm, the missing values in each sample are "imputed" by utilising the mean value of the closest neighbours that can be located in the training set. Two samples are considered comparable if the characteristics that neither one of them is lacking are comparable. 2,704 missing values are imputed using this algorithm. Signal missing label matrix graph after column drop is represented in a signal missing label matrix graph is represented in Figure 7, where Figure 7(a) represents a graph after column drop and Figure 7(b) represents a graph after the KNNImputer transformer. Graph of semi-conductor fault checkup data is given in Figure 8.

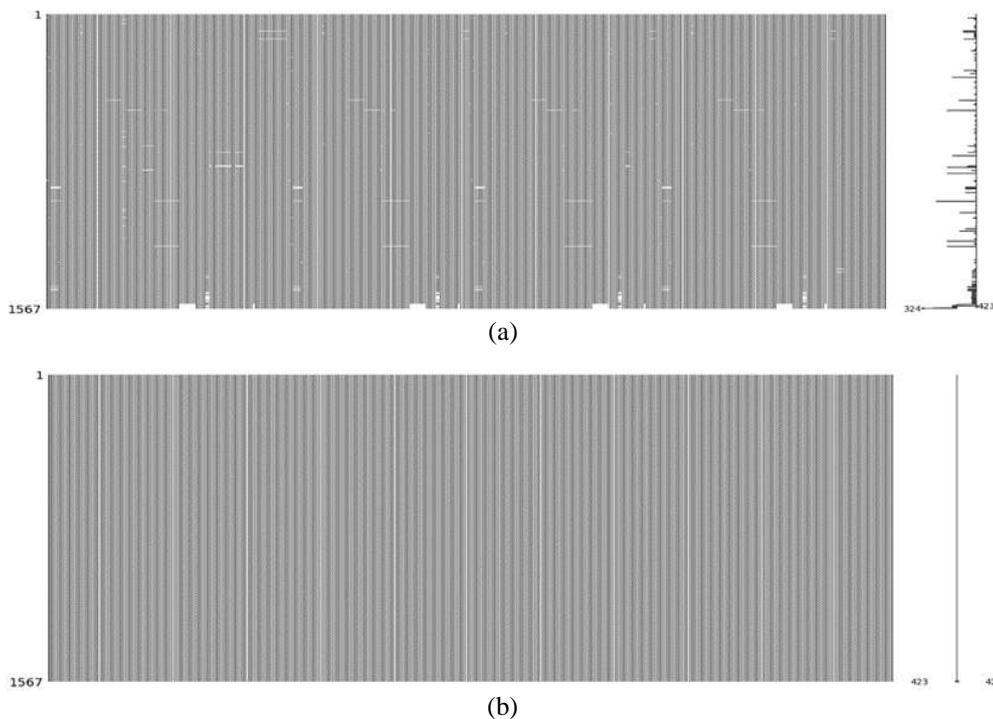


Figure 7. Signal missing label matrix graph (a) signal missing label matrix graph after column drop and (b) signal missing label matrix graph after KNNImputer transformer

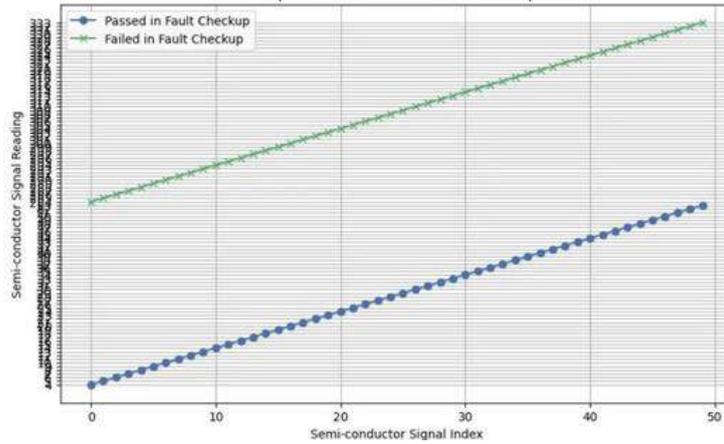


Figure 8. Graph of semi-conductor fault checkup data

**4.2. Feature selection and extraction**

Outlier identification using variance inflation factor is the initial feature selection algorithm that is used in the proposed research. The variance inflation factor is a metric which defines how much the behavior (variance) of an independent variable is modified or enlarged as a result of its correlation or interaction with the behavior of the other independent variables. In other words, the variance inflation factor measures the degree to which one independent variable affects or correlates with the behavior of other independent variables. A simple measurement of how much a variable is contributing in the regression to the standard error can be obtained through the use of variance inflation factors. It can be seen that after applying VIF algorithm, there are a number of columns containing outliers. Outlier high and low for selected columns is represented in Figure 9. The columns containing high outliers are dropped. Shape after dropping outliers is (1,567, 237). Outlier high and low after dropping is represented in Figure 10.

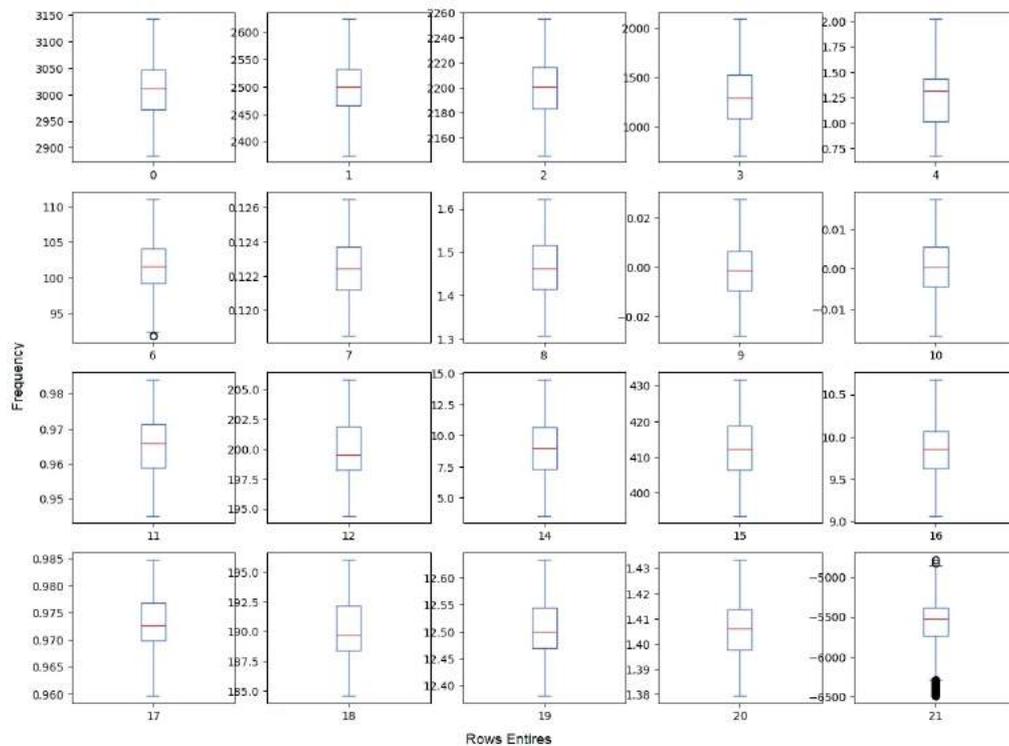


Figure 9. Outlier high and low for selected columns

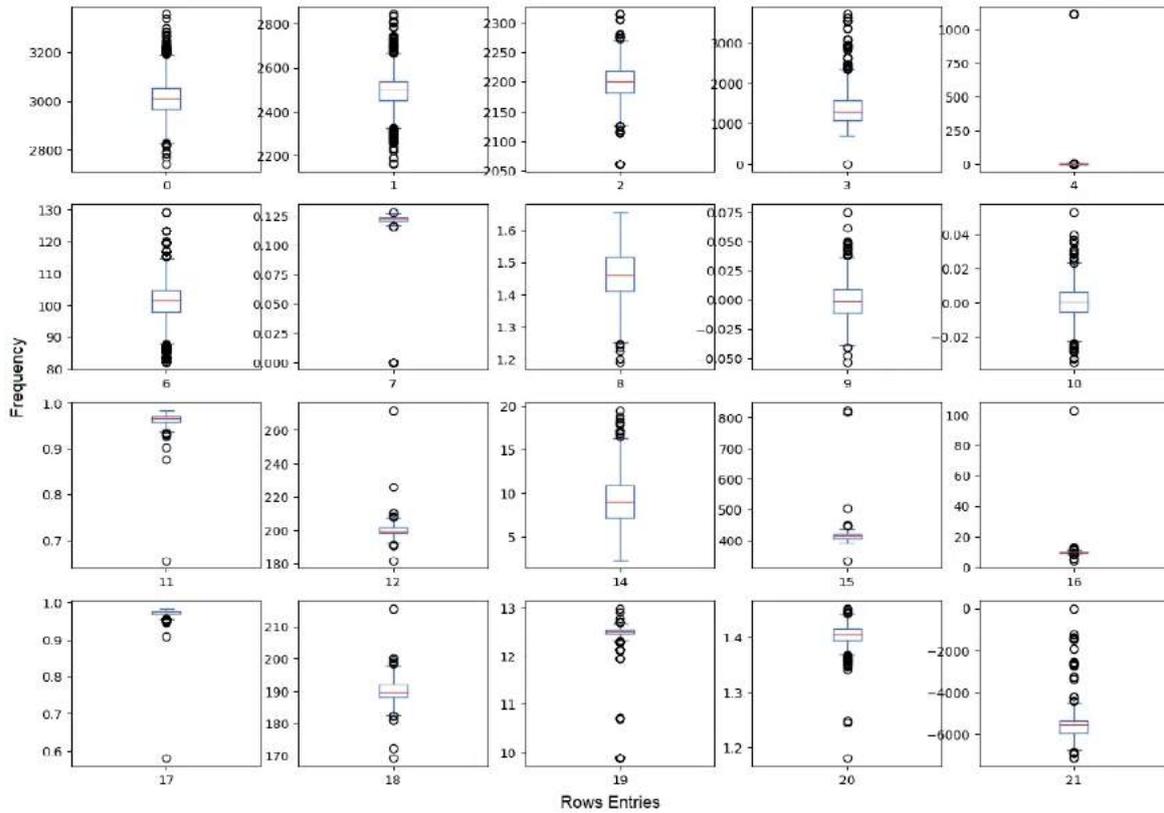


Figure 10. Outlier high and low after dropping

**4.3. Standardization and distribution**

MinMax scalar is used for standardization. The MinMaxscaler is a specific kind of scaler that adjusts the minimum and maximum values so that they are, respectively, 0 and 1. Whereas the StandardScaler adjusts all of the values between the minimum and maximum thresholds such that they fit inside the specified range of values from minimum to maximum. Most of the columns show a normal distribution which looks good, but few columns show a bimodal distribution. Pass fail yield graph of both distribution is represented in Figure 11.

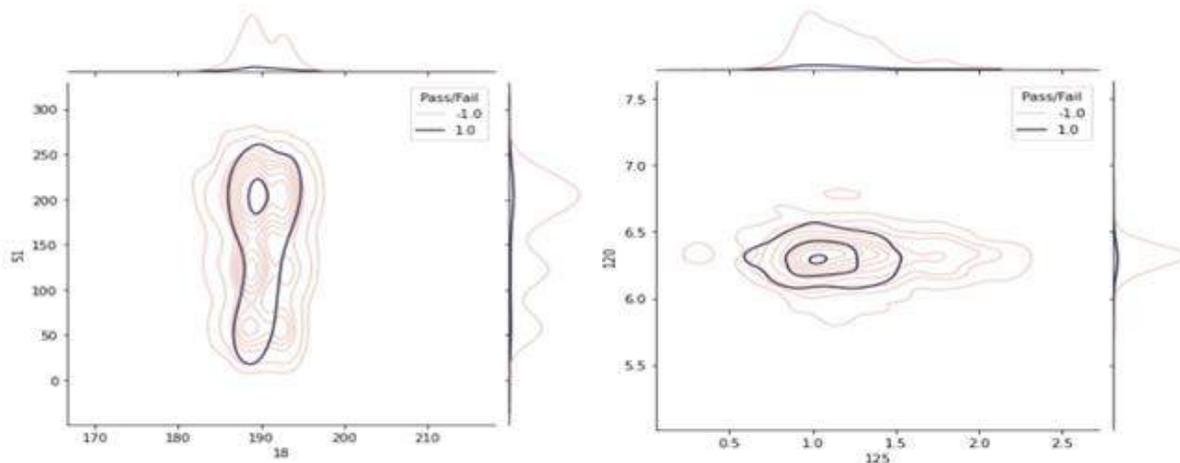


Figure 11. Pass fail yield graph of both distribution

The higher intensity curves represent the pass yield, and the lower intensity graph represents the fail yield. The trajectory of axis is represented as graph outside the box. Raw data based clustering graph of pass-fail yield is represented in Figure 12. Pass/fail yield clustering graph of data after standardization is represented in Figure 13.

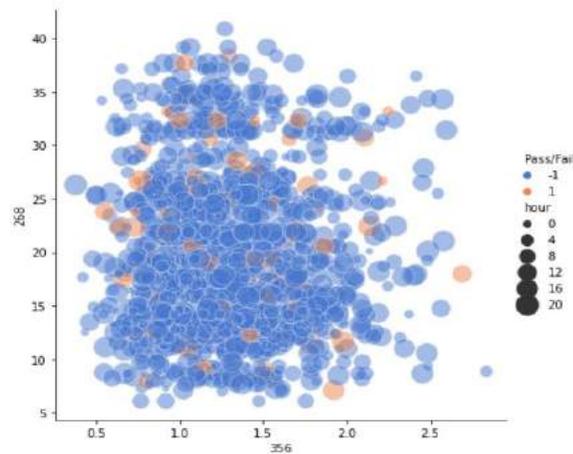


Figure 12. Raw data based clustering graph of pass fail yield

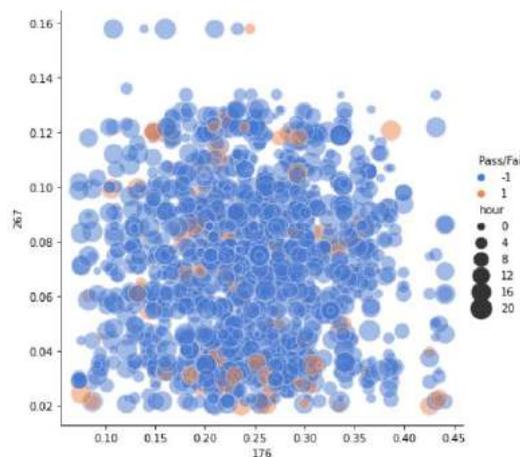


Figure 13. Pass/fail yield clustering graph of data after standardization

**4.4. Analysis using principal component analysis and select k best algorithm**

PCA is done on the dataset after standardization. PCA is a technique for extracting essential variables (in the form of components) from a vast set of variables that are made available in a data set. This is accomplished through the use of principal components. It does this by taking a projection of unimportant dimensions from a high-dimensional data set and then using that as a basis for extracting a low-dimensional set of features with the goal of obtaining as much information as feasible. The visualisation process also becomes significantly more relevant when fewer variables are obtained while the amount of information that is lost is minimised. When working with data that has three or more dimensions, PCA proves to be more beneficial. PCA variance graph is represented in Figure 14. PCA component transformation is the number of dimensions needed to capture 90% of variance is 143. Shape after PCA is (1,567, 143). Correlation matrix with and without PCA is represented in Figure 15.

Select K best is also used for component analysis. It is another algorithm that is used for component analysis. Matrix shape created by select K best is (1,567,50). Select K best variance graph is represented in Figure 16. The select K best component transformation is ((-1: 1463, 1: 104)).

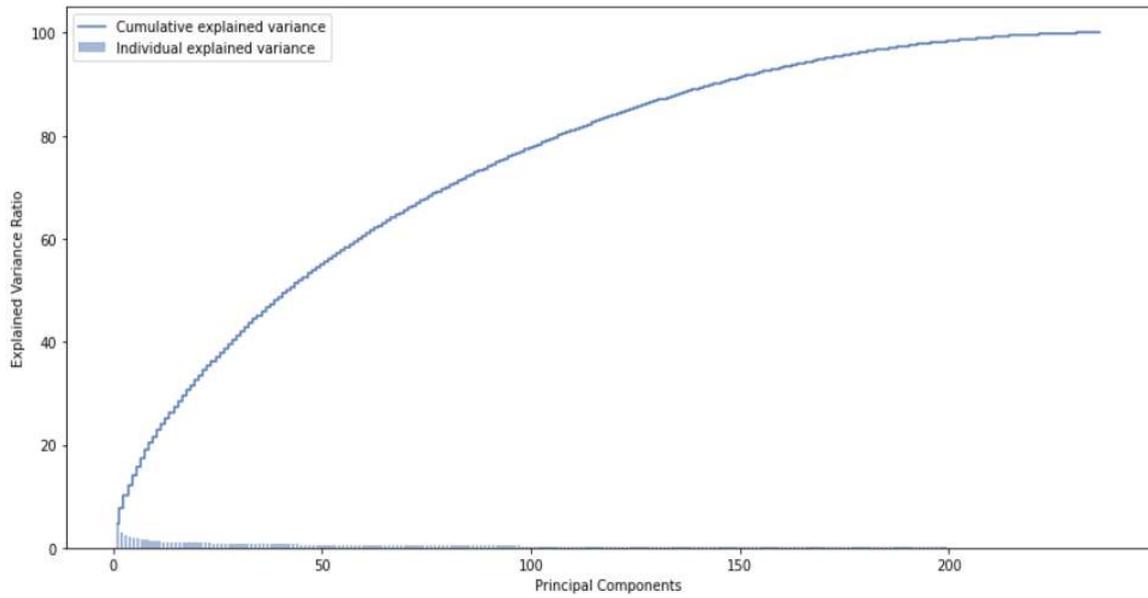


Figure 14. PCA variance graph

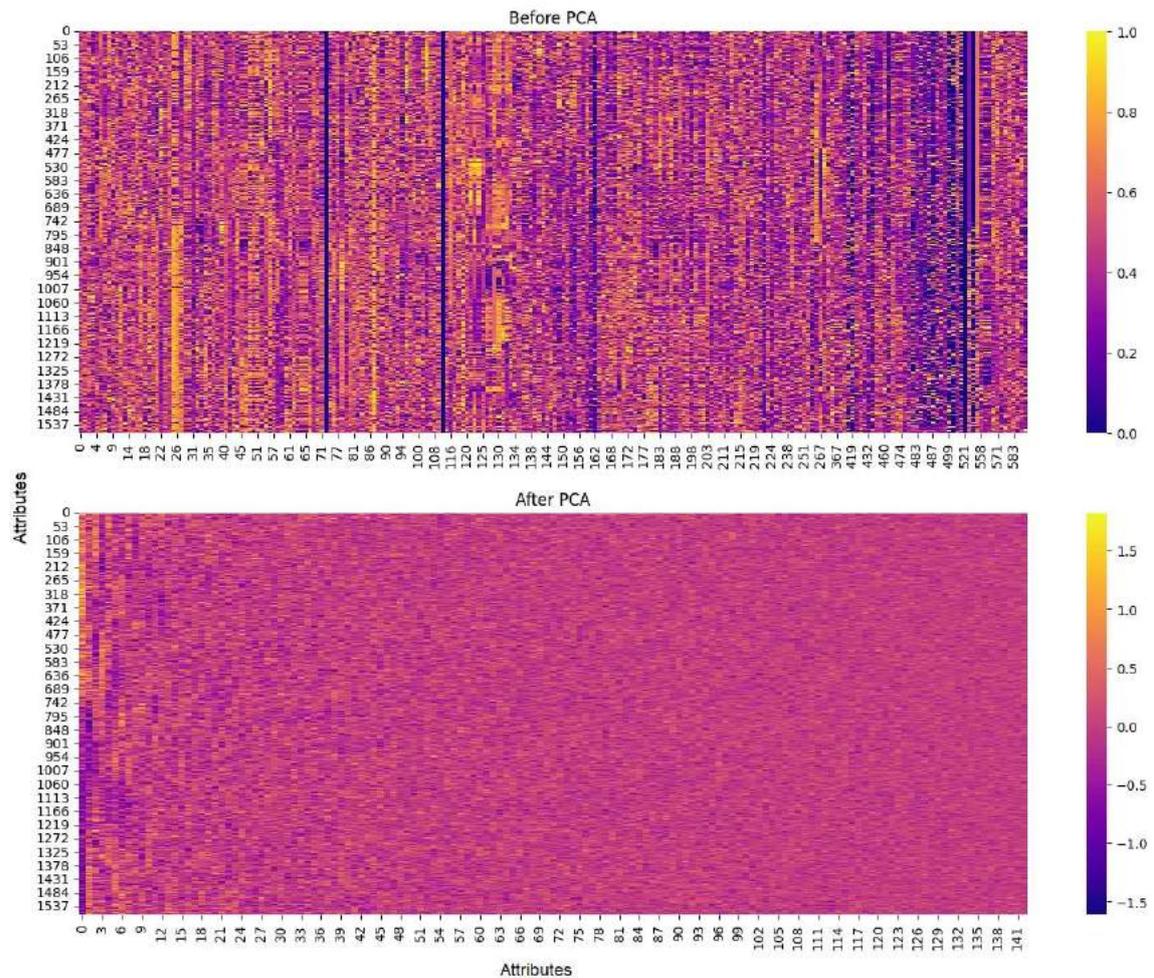


Figure 15. Correlation matrix with and without PCA

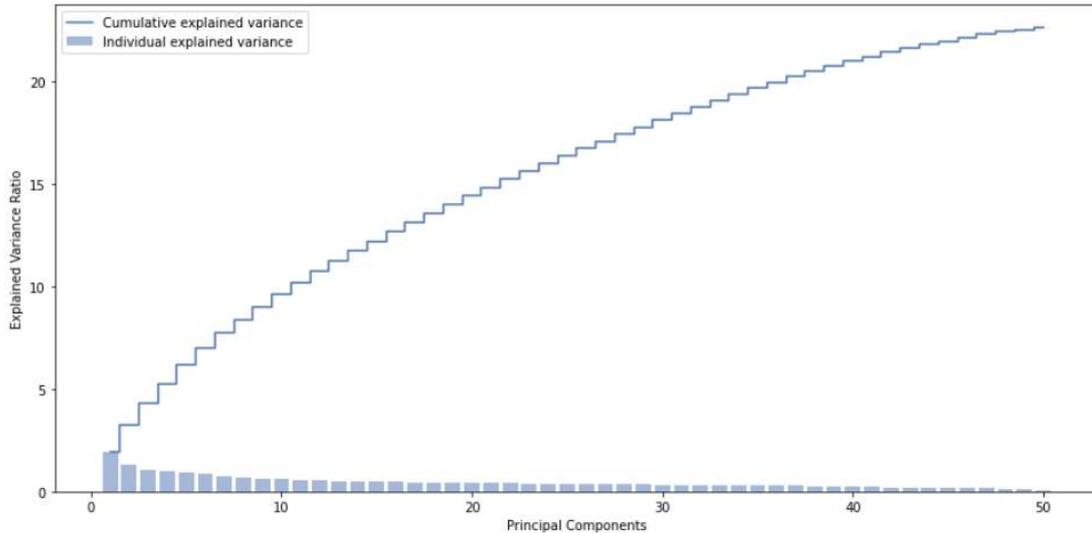


Figure 16. Select K best variance graph

**4.5. Inter band redundancy analysis**

The redundancy analysis is performed on the final dataset using IBRA. The IBRA initial data is represented in Figure 17. Correlated data is represented in Figure 18. Next cross decomposition is performed for IBRA. IBRA cross decomposition correlation score graph is represented in Figure 19.

	0	1	2	3	4	5	6	7	8	9	...	133	134	135	136
0	0.897612	-0.256396	0.526373	0.196095	0.064767	0.461568	0.023081	0.128638	0.469676	-0.132360	...	0.144517	-0.401555	-0.257179	0.011108
1	0.319438	-0.425457	0.441982	0.036232	0.627701	-0.018057	-0.285748	0.266533	0.105831	0.115399	...	0.083015	-0.052544	0.007346	0.064581
2	0.065730	0.522600	0.506394	-0.059330	0.532804	-0.230805	0.019155	-0.317627	0.457494	-0.084984	...	-0.000494	0.020786	0.404664	0.043515
3	0.278178	0.377714	0.561972	0.012374	0.493327	0.016169	-0.108919	-0.186663	-0.459912	0.359271	...	0.106468	-0.183325	0.128352	0.010099
4	0.697863	-0.014978	0.354104	-0.142276	0.432032	0.089229	-0.789594	0.050809	0.210124	0.514384	...	-0.055778	-0.107763	-0.074057	-0.121408

Figure 17. IBRA initial data

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0.137680	0.204425	-0.020680	0.209900	-0.089708	0.284158	-0.209468	0.108749	0.114046	-0.381274	-0.322914	-0.459245	0.332582	-0.608175
1	-0.039240	0.091576	0.269525	0.120701	0.349271	0.235725	0.310679	-0.068658	-0.265969	-0.514304	-0.253314	-0.106354	0.171849	-0.405519
2	0.260663	-0.086882	-0.030394	-0.193629	-0.342639	0.031782	-0.179977	0.008338	0.049807	-0.061642	0.189135	0.417860	-0.149440	-0.021034
3	0.232788	0.280343	0.254479	0.055008	-0.120949	0.177058	0.100872	-0.083662	0.171758	-0.324372	-0.137795	-0.154135	-0.070307	0.144150
4	-0.239404	-0.607853	0.076398	0.457118	0.030871	0.372783	-0.250464	-0.077037	-0.545530	-0.107200	-0.124765	-0.030009	-0.086106	0.257921

Figure 18. Correlated data

**4.6. Greedy spectral selection**

Additional feature selection and extraction is performed by the GSS algorithm. selected bands after GSS feature selection and extraction is [1, 0, 2, 3, 4, 5, 6, 7, 8, 9]. Smote over sampling is represented by (-1: 1,023, 1:1,023). Dataset after applying GSS feature selection and extraction is given in Figure 20.

**4.7. Anomaly detection using k-means clustering**

The K-means clustering, DBScan clustering and mean shift clustering techniques are considered for anomaly detection and it is seen that K-means clustering has better cluster labels selection for the dataset. The scores for Silhouette, Davies\_bouldin, and Calinski Harabasz are given in Table 1. The K-mean cluster creation is done based on local outlier factor (LOF) and outlier clusters are trained using isolation forest. The anomaly graph is given in Figure 21.

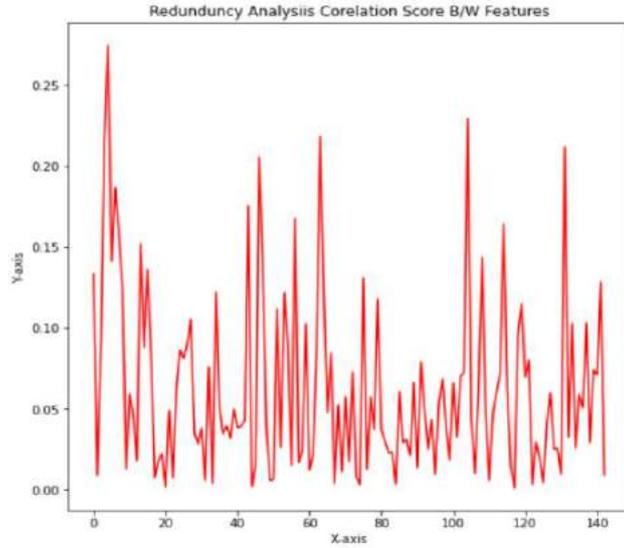


Figure 19. IBRA cross decomposition correlation score graph

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0.137680	0.204425	-0.020680	0.209900	-0.089708	0.284158	-0.209468	0.108749	0.114046	-0.381274	-0.322914	-0.459245	0.332582	-0.608175
1	-0.039240	0.091576	0.269525	0.120701	0.349271	0.235725	0.310679	-0.068658	-0.265969	-0.514304	-0.253314	-0.106354	0.171849	-0.405519
2	0.260663	-0.086882	-0.030394	-0.193629	-0.342639	0.031782	-0.179977	0.008338	0.049807	-0.061642	0.189135	0.417860	-0.149440	-0.021034
3	0.232788	0.280343	0.254479	0.055008	-0.120949	0.177058	0.100872	-0.083662	0.171758	-0.324372	-0.137795	-0.154135	-0.070307	0.144150
4	-0.239404	-0.607853	0.076398	0.457118	0.030871	0.372783	-0.250464	-0.077037	-0.545530	-0.107200	-0.124765	-0.030009	-0.086106	0.257921

Figure 20. Dataset after feature selection extraction

Table 1. Comparison of clustering algorithms

Clustering technique	Silhouette	Davies bouldin	Calinski harabasz
K-Means	0.408	2.958	147.670
DBScan	0.396	1.797	7.689
Mean shift	0.298	1.479	25.940

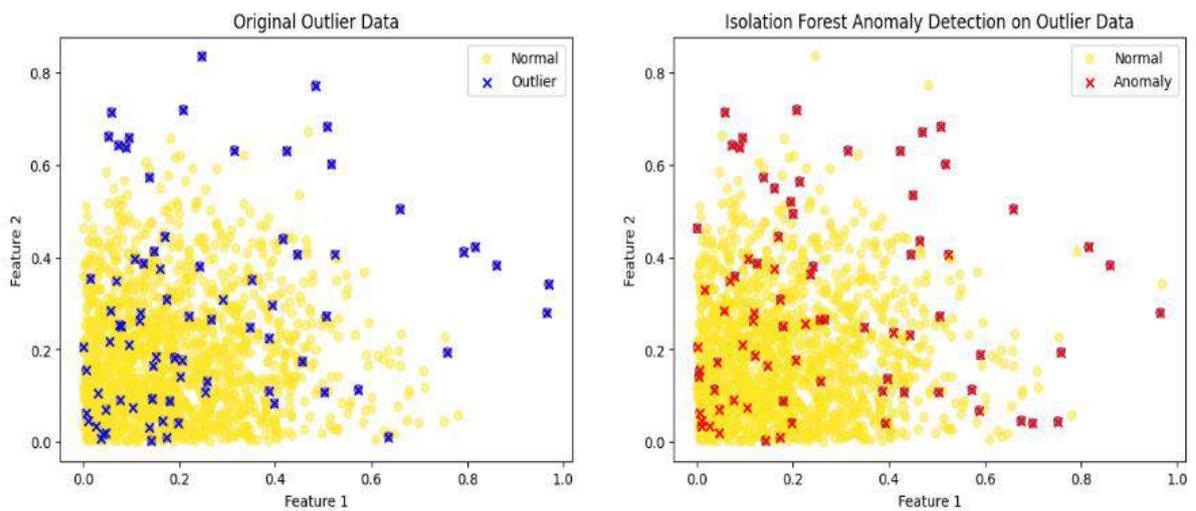


Figure 21. Anomaly graph

**4.8. Convolutional neural network with bacterial foraging optimization**

The dataset is trained and tested using convolutional neural network for predicting and analysing its performance. The dataset is split into train and test data in the ratio 70% to 30%. ((1,096, 10), (471, 10), (1,096,), (471,)) The CNN accuracy score is 85.56%. Bacterial foraging optimization is used for hyperparameter tuning of the data. “Hyperparameter tuning” refers to the process of determining the optimal values for the learning algorithm's hyperparameters so that the enhanced version of the algorithm can be applied to any data set. This must be done while maintaining the flexibility to apply the algorithm to any data set. This specific combination of hyperparameters helps to optimize the performance of the model by lowering the overall size of a loss function that has been defined in the past. As a direct consequence of this, the model produces findings that are more precise and that contain less errors. For optimization using BFO, the selected number of points is 20. The fitness function and hyperparameters are represented in Figure 22.

```

Generation 1
15/15 [=====] - 0s 2ms/step
(471,)
15/15 [=====] - 0s 936us/step
(471,)
15/15 [=====] - 0s 827us/step
(471,)
15/15 [=====] - 0s 960us/step
(471,)
15/15 [=====] - 0s 2ms/step
(471,)
15/15 [=====] - 0s 2ms/step
(471,)
15/15 [=====] - 0s 1ms/step
(471,)
15/15 [=====] - 0s 2ms/step
(471,)
15/15 [=====] - 0s 1ms/step
(471,)
15/15 [=====] - 0s 2ms/step
(471,)
15/15 [=====] - 0s 1ms/step
(471,)
15/15 [=====] - 0s 2ms/step
(471,)
15/15 [=====] - 0s 1ms/step
(471,)
15/15 [=====] - 0s 2ms/step
(471,)
15/15 [=====] - 0s 1ms/step
(471,)
15/15 [=====] - 0s 2ms/step
(471,)
15/15 [=====] - 0s 2ms/step
(471,)
15/15 [=====] - 0s 1ms/step
(471,)
15/15 [=====] - 0s 2ms/step
(471,)
15/15 [=====] - 0s 1ms/step
(471,)
Best fitness: 0.06581740976645435
Best params: {'filters': 63.89424951841391, 'kernel_size': 4.939668843421117, 'pool_size': 3.1333519217359957, 'dense_units': 1
5.926289416678074, 'learning_rate': -0.10510521502040622}
Generation 2
15/15 [=====] - 0s 934us/step
(471,)
15/15 [=====] - 0s 2ms/step
(471,)

```

Figure 22. Fitness function and hyperparameters

Fitness function and hyperparameters-training CNN using BFO hyperparameter Tuning for 10 generation. BFO is calculating optimum in each iteration and providing it to CNN and CNN is running all iteration for each parameters set in hyperparameters for 10 generation. The best fitness function is 0.668. The CNN model’s training loss graph is given in Figure 23. The area under the curve (AUC) or receiver operating characteristics (ROC) scores of the proposed metTablehod is ROC score is 0.971 and AUC score is 0.813. The other scores are given in Table 2.

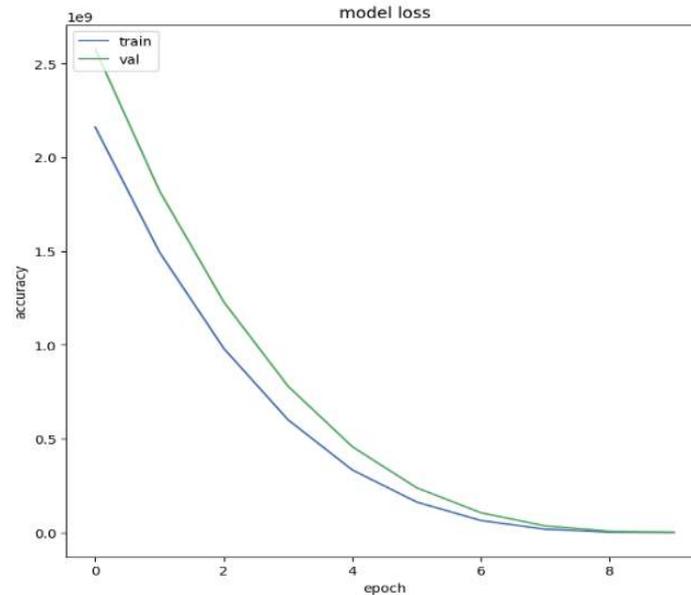


Figure 23. CNN model's training loss graph

Table 2. Performance scores of the proposed model

Metrics	Values
Mean squared error	0.460
Mean absolute error	0.106
R2 score	0.136
Max. error	2.0
Mean pinball loss	0.0106
Root mean squared error	0.678
Mean absolute percentage error	10.615

#### 4.9. Comparison of the proposed hybrid algorithm with other state of the art algorithms in the literature

The proposed hybrid algorithm is compared with other state of the art algorithms in literature. The algorithms that are used for comparison are Naive Bayes, K neighbours, support vector machine (SVM), logistic regression, decision tree, and random forest. It can be seen that the proposed method gives the highest accuracy with 94.69%. The accuracy of the other algorithms is lower than the proposed algorithm. Table 3 gives the comparison of the proposed algorithm with similar algorithms in the literature.

Table 3. Comparison of proposed algorithm with similar algorithms in the literature

Algorithm/Metric	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
CNN & BFO	94.69	77.67	97.15	84.17
Naive bayes	70	63	84	72
KNeighbours	72	56	97	71
SVM	73	66	87	75
Logistic regress	57	56	61	58
Decision tree	57	56	65	60
Random forest	70	63	84	72

## 5. CONCLUSION

This study was mostly about suggesting a filter-based (IBRA) and a hybrid (IBRA-GSS) band selection approach in the proposed research, which is about sorting yield data in the semiconductor industry. Our tests show that our IBRA method can get rid of a lot of the spectral bands while still doing a good job of classifying data, just like when we used the whole spectrum. As a result, IBRA is able to successfully eliminate superfluous spectral bands while preserving the spectral information that is most important for a particular categorization endeavor.

In addition, our IBRA-GSS technique was offered as a hybrid selection method that enables the selection of a predetermined number of spectral bands  $k$ . This was done in order to facilitate the selection process. We also made the observation that the IBRA step can have more than one VIF threshold, and this can lead to the same or a comparable selection of bands, both of which were discovered to be equally effective. As the size of the dataset decreased, it was observed that this trend remained the same. In the context of the proposed research, the convolutional neural network is applied to the task of performance evaluation, and the bacterial foraging optimization algorithm is applied to the enhancement of the performance matrices. The paper's proposed method for mining HSI's spatial information has not been implemented before. As a result, our future efforts will center on finding ways to include geographical information in the feature selection process.

## ACKNOWLEDGMENT

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2023/01/ 25865).

## REFERENCES

- [1] B. Scherrer, J. Sheppard, P. Jha, and J. A. Shaw, "Hyperspectral imaging and neural networks to classify herbicide-resistant weeds," *Journal of Applied Remote Sensing*, vol. 13, no. 04, p. 1, Nov. 2019, doi: 10.1117/1.JRS.13.044516.
- [2] G. C. Davenport, "Remote sensing applications in forensic investigations," *Historical Archaeology*, vol. 35, no. 1, pp. 87–100, Mar. 2001, doi: 10.1007/BF03374530.
- [3] S. A. Mathews, "Design and fabrication of a low-cost, multispectral imaging system," *Applied Optics*, vol. 47, no. 28, p. F71, Oct. 2008, doi: 10.1364/AO.47.000F71.
- [4] C. Yang, "A high-resolution airborne four-camera imaging system for agricultural remote sensing," *Computers and Electronics in Agriculture*, vol. 88, pp. 13–24, Oct. 2012, doi: 10.1016/j.compag.2012.07.003.
- [5] D. Wu and D.-W. Sun, "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part I: Fundamentals," *Innovative Food Science & Emerging Technologies*, vol. 19, pp. 1–14, Jul. 2013, doi: 10.1016/j.ifset.2013.04.014.
- [6] A. F. H. Goetz, "Three decades of hyperspectral remote sensing of the earth: a personal view," *Remote Sensing of Environment*, vol. 113, pp. S5–S16, Sep. 2009, doi: 10.1016/j.rse.2007.12.014.
- [7] C. M. Lee *et al.*, "An introduction to the NASA hyperspectral infrared imager (HypIRI) mission and preparatory activities," *Remote Sensing of Environment*, vol. 167, pp. 6–19, Sep. 2015, doi: 10.1016/j.rse.2015.06.012.
- [8] X. Guo, H. Zhang, Z. Wu, J. Zhao, and Z. Zhang, "Comparison and evaluation of annual NDVI time series in China derived from the NOAA AVHRR LTDR and terra MODIS MOD13C1 products," *Sensors*, vol. 17, no. 6, p. 1298, Jun. 2017, doi: 10.3390/s17061298.
- [9] M. González-Betancourt and Z. L. Mayorga-Ruiz, "Normalized difference vegetation index for rice management in el Espinal, Colombia," *DYNA (Colombia)*, vol. 85, no. 205, pp. 47–56, 2018, doi: 10.15446/dyna.v85n205.69516.
- [10] A. Martínez-Uso, Martínez-Uso, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007, doi: 10.1109/TGRS.2007.904951.
- [11] O. S. Sonbul and M. Rashid, "Algorithms and techniques for the structural health monitoring of bridges: systematic literature review," *Sensors*, vol. 23, no. 9, p. 4230, Apr. 2023, doi: 10.3390/s23094230.
- [12] L. Deng, Z. Mao, X. Li, Z. Hu, F. Duan, and Y. Yan, "UAV-based multispectral remote sensing for precision agriculture: a comparison between different cameras," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 146, pp. 124–136, Dec. 2018, doi: 10.1016/j.isprsjprs.2018.09.008.
- [13] L. P. Aggarwal and F. A. Papay, "Applications of multispectral and hyperspectral imaging in dermatology," *Experimental Dermatology*, Jun. 2022, doi: 10.1111/exd.14624.
- [14] G. Morales, J. Sheppard, R. Logan, and J. Shaw, "Hyperspectral band selection for multispectral image classification with convolutional networks," in *2021 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2021, pp. 1–8, doi: 10.1109/IJCNN52387.2021.9533700.
- [15] J. Darbon and S. Osher, "Algorithms for overcoming the curse of dimensionality for certain Hamilton-Jacobi equations arising in control theory and elsewhere," *Research in the Mathematical Sciences*, vol. 3, no. 1, p. 19, Dec. 2016, doi: 10.1186/s40687-016-0068-7.
- [16] W. Sun and Q. Du, "Hyperspectral band selection: a review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 118–139, Jun. 2019, doi: 10.1109/MGRS.2019.2911100.
- [17] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968, doi: 10.1109/TIT.1968.1054102.
- [18] M. Hasheminejad, "Optimized kernel nonparametric weighted feature extraction for hyperspectral image classification," *Journal of Information Systems and Telecommunication (JIST)*, vol. 10, no. 38, pp. 111–119, Apr. 2022, doi: 10.52547/jist.16105.10.38.111.
- [19] S. Prasad and L. M. Bruce, "Limitations of principal components analysis for hyperspectral target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 625–629, Oct. 2008, doi: 10.1109/LGRS.2008.2001282.
- [20] C. Jayaprakash, B. B. Damodaran, S. Viswanathan, and K. P. Soman, "Randomized independent component analysis and linear discriminant analysis dimensionality reduction methods for hyperspectral image classification," *Journal of Applied Remote Sensing*, vol. 14, no. 03, p. 1, Jul. 2020, doi: 10.1117/1.JRS.14.036507.
- [21] M. Fordellone, A. Bellincontro, and F. Mencarelli, "Partial least squares discriminant analysis: a dimensionality reduction method to classify hyperspectral data," *Statistica Applicata*, vol. 31, no. 2, pp. 181–200, 2019, doi: 10.26398/IJAS.0031-010.
- [22] J. Lapajne, M. Knapič, and U. Žibrat, "Comparison of selected dimensionality reduction methods for detection of root-knot nematode infestations in potato tubers using hyperspectral imaging," *Sensors*, vol. 22, no. 1, p. 367, Jan. 2022, doi: 10.3390/s22010367.

- [23] Masashi Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.
- [24] D. Hong, N. Yokoya, and X. X. Zhu, "Learning a robust local manifold representation for hyperspectral dimensionality reduction," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 6, pp. 2960–2975, Jun. 2017, doi: 10.1109/JSTARS.2017.2682189.
- [25] J. Ren, R. Wang, G. Liu, R. Feng, Y. Wang, and W. Wu, "Partitioned relief-F method for dimensionality reduction of hyperspectral images," *Remote Sensing*, vol. 12, no. 7, p. 1104, Mar. 2020, doi: 10.3390/rs12071104.
- [26] M. Parvathi, N. Vasantha, and K. S. Parasad, "Modified march C - algorithm for embedded memory testing," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 2, no. 5, Oct. 2012, doi: 10.11591/ijece.v2i5.1587.
- [27] "UCI SECOM Dataset," Semiconductor manufacturing process dataset. [Online]. Available: <https://www.kaggle.com/datasets/paresh2047/uci-semcom>.

## BIOGRAPHIES OF AUTHORS



**Dr. Mohammed Altaf Ahmed**     was born in India. He received his Ph.D. in Engineering from GITAM University, Vishakhapatnam, India, for the thesis entitled "Design of a built-in self-test controller using the march algorithm for fault diagnosis in embedded memories" in the year 2018. He received the Best Thesis Award and the Gold Medal for his Ph.D. He completed a master's in engineering and technology with a specialization in embedded systems in 2008 from Jawaharlal Nehru Technological University, Hyderabad, India. He received his bachelor's in engineering (Electronics and Communications) in 2000 from Swami Ramanand Teerth Marathwada University, Nanded, India. He is presently working in the College of Computer Engineering and Sciences at Prince Sattam bin Abdulaziz University (PSAU), Kingdom of Saudi Arabia, as an assistant professor in the Computer Engineering Department. Additionally, he has been working in the Quality and Development Unit, Computer Engineering Department. He is presently focusing his research on VLSI design, embedded systems, and IoT devices. He has successfully implemented several research grant projects in the same area and published research articles in leading academic journals indexed in Scopus and Science. He can be contacted at email: [m.altaf@psau.edu.sa](mailto:m.altaf@psau.edu.sa).



**Dr. Suleman Alnatheer**     is an assistant professor in computer engineering at Prince Sattam bin Abdulaziz University. In this role, he conducts research on the topic of statistical learning theory that can be applied to next-generation wireless communication technologies. In addition, he teaches undergraduate courses in various electrical and computer engineering topics. Prior to joining Prince Sattam bin Abdulaziz University, he worked as a lecturer in the College of Telecommunication and Information, where he was lucky enough to work with amazing people. He holds a Ph.D. from Stevens Institute of Technology and an M.Sc. and B.Sc. from George Washington University. He can be contacted at email: [s.alnatheer@psau.edu.sa](mailto:s.alnatheer@psau.edu.sa).