IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 April 2023, accepted 29 April 2023, date of publication 8 May 2023, date of current version 17 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3274490

RESEARCH ARTICLE

Evolutionary Approach for Interpretable Feature Selection Algorithm in Manufacturing Industry

SANGHOUN OH[1] AND CHANG WOOK AHN[2], (Member, IEEE)
[1]Department of Computer Science, Korea National Open University, Seoul 03087, Republic of Korea
[2]AI Graduate School, Gwangju Institute of Science and Technology, Gwangju 61005, Republic of Korea

Corresponding author: Chang Wook Ahn (cwan@gist.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (No. NRF-2021R1A2C3013687), and the Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government (MSIT)(No. 2019-0-01842, Artificial Intelligence Graduate School Program (GIST)).

ABSTRACT Feature selection techniques in prediction play a role in manufacturing industries of late. However, it is very challenging to achieve an optimal subset of features as well as interpretable relationship among features due to computation complexity and variable diversity. In order to address those difficulties, this paper presents a novel evolutionary approach for feature selection algorithm to improve the effectiveness of existing meta-heuristic approaches. In other words, their optimal combinations with minimal difference between prediction and actual values can be achieved by applying an estimation of distribution algorithms (i.e., extended compact genetic algorithm) on the collected candidate feature sets. The approach discovers a less complicated and more closely related probabilistic-model structure on population space in each generation, thereby encouraging the comprehension power of feature selection results. We tested our method on six real-world data sets from manufacturing industries (open to the public). It demonstrated that higher interpretability on features selection results is achieved in comparison with well-known methods.

INDEX TERMS Evolutionary-based approach, feature selection, extended compact genetic algorithm, manufacturing industries.

I. INTRODUCTION

In the case of real manufacturing problems, it is a very complex and important task to detect process abnormalities in advance through real-time processing of large amounts of data collected from the manufacturing environment [1], [3], [4], [5], [6]. Those data sets being processed in this way consist of several properties and features. Typically, a series of processes to obtain useful information is performed by utilizing all features defined in the given data set; however, in some features, the target lacks relevance and the performance may be degraded in modeling for obtaining useful information due to duplication [1]. To apply machine learning techniques into real-time large-scale data collected in the manufacturing industry, selecting key features is a significant task, though it is one of the most difficult tasks [1], [5] [7]. In the given data, if a total of n features is included, a total of 2^n subsets should be possible, and the best subset must be selected. Here, if n is a large number, it will become difficult to do a performance evaluation in this problem. To effectively resolve these problems, a variety of methodologies has been proposed [1]. Firstly, conventional searching algorithms (i.e., exhaustive search, greedy search, and random search) are applied into feature selections for finding the best subset. In the case of this method, it is quite complicated to search for optimal features due to early convergence, enormous complexity, and high computational cost. To overcome this issue, other feature selection methods on meta-heuristic algorithms have been recently proposed since they are the most efficient and effective technique and allows you to detect the relevant subset of features and to maintain model accuracy at the same time [1], [3], [5], [7], [8]. In this study, we describe a novel

The associate editor coordinating the review of this manuscript and approving it for publication was Li Zhang.

interpretable feature selection method by using a well-known linkage learning in genetic algorithm (i.e., extended compact genetic algorithm: ECGA).

II. LITERATURE SURVEY

A. FEATURE SELECTION

It aims to efficiently remove inappropriate, irrelevant, or unnecessary features-in other words optimal features are extracted from a given dataset. However, it is one of the most important and difficult problems as employing machine learning techniques to large-scale data collected as real-world applications such as bioinformatics for finding the best gene in a candidate gene and text mining for finding the best term word or phrase [1], [5], [7].

Mathematically, the statement of feature selection can be specified in following way. Suppose a given dataset consists of d features such as. Then the process of this selection problem is to extract the optimal subset of features composed of n number of features where $n < d$. To extract the best features combinations, several feature selection methods have following way been developing as classifying three categories (i.e., filter, wrapper, and embedded methods). Firstly, the filter method independently operates the learning or classification algorithm - in other words it is totally dependent on the given data [1]. Next, the wrapper technique always includes machine learning algorithms and extracts the optimal subset of features through interaction [1], [9]. This technique provides more accurate results than the filter method, though it is much more computationally expensive than the filter method. Lastly, the embedded method mixes filter and wrapper methods. This study proposes a new wrapper method on meta-heuristic of estimation of distribution algorithms (EDA); that is, extended compact genetic algorithm (ECGA).

B. META-HEURISTIC ALGORITHMS

The aim of meta-heuristic algorithms working based on probability is to obtain a close optimal solution in each problem as avoiding local optima by randomly generating a group of solution candidates with simplicity, flexibility, and the ability. For achieving the optimum, they play key roles of exploration and exploitation [1], [3]. In the exploration, its algorithm thoroughly investigates the promising search space and utilizes it for local search of the promising regions found in the exploitation stage. There are several adjusting applications such as electrical engineering (e.g., power generation), industrial engineering (e.g., work scheduling), civil engineering (e.g., design architecture), and telecommunications engineering (e.g., radar design, networking) [1], [2]. Especially, in black-box models in meta-heuristics, mining that human can't understand how variables are being combined to find optimum because they are directly established from data by an algorithm with complicated functions of the variables [3], [10].

There are two major categories of meta-heuristic algorithms. One type is single solution based meta-heuristic algorithms initiating the optimization process with one solution updated during iteration. They may end up stuck in a local optimal, and you won't even thoroughly explore the search space. The other type is multiple solution based meta-heuristic algorithms initiate the optimization process as generating a population of solutions updated by iterations. They are useful for preventing local optimum, as multiple solutions support each other and navigate the search space well. They also have qualities that make them a promising part of the search space, so they are used to solve most real-world problems (especially, manufacturing industries).

Furthermore, meta-heuristic algorithms are classified with four categories according to their behaviors: evolution-based, swarm intelligence-based, physics-based, and human-related algorithms [1]. The first evolution-based algorithms begin to randomly generated population of solutions inspired by Darwin natural evolution. In those algorithms, a new solution is generated through the two main genetic operators (i.e., crossover and mutation) for achieving the best solution through repetition of this process [11], [12]. There are several state-of-the-art algorithms: genetic algorithm, evolution strategy, genetic programming, tabu search, and differential evolution. The second swarm-intelligence based algorithms were developed based on social behavior patterns of insects, animals, and birds as traversing search spaces and best locations. There are several representative algorithms (i.e., ant colony optimization that applies pheromone-based commutation of biological ants, honey bee swarm optimization algorithm that utilizes bee behavior mechanisms, and monkey optimization that depicts patterns in a herd of monkeys) [13], [14]. The third physics-based algorithms such as simulated annealing and harmony search are affiliated by the principals of physics in the universe [1]. The last human behavior-based algorithms are entirely inspired by behavioralism. Everyone has a way of affecting his or her performance [1].

III. EVOLUTIONARY APPROACH FOR INTERPRETABLE FEATURE SELECTION ALGORITHM

In general, the-state-of-art meta-heuristic algorithms could extract the optimal combination of features by designing the given problems with respect to each characteristic. Despite this strength, their algorithms are much tougher to perceive the importance of each feature and to understand how the different features interact in predictions or classifications [1], [5], [7]. To effectively resolve this limitation, this paper represents a new interpretable feature selection method developed by a linkage learning based evolutionary algorithm that is extended compact genetic algorithm proposed by Harik. It is equivalent to linkage learning as selecting of a good probability distribution measured by quantification using minimum description length (MDL) model [15], [16], [16], [17]. A key concept of this model is that a simple distribution is better than a complex distribution when all things are equal. To give rise to an optimal probability distribution, the MDL restriction disciplines inaccuracy and complexity of models. Hence, MDL regulation creates the problem of finding a

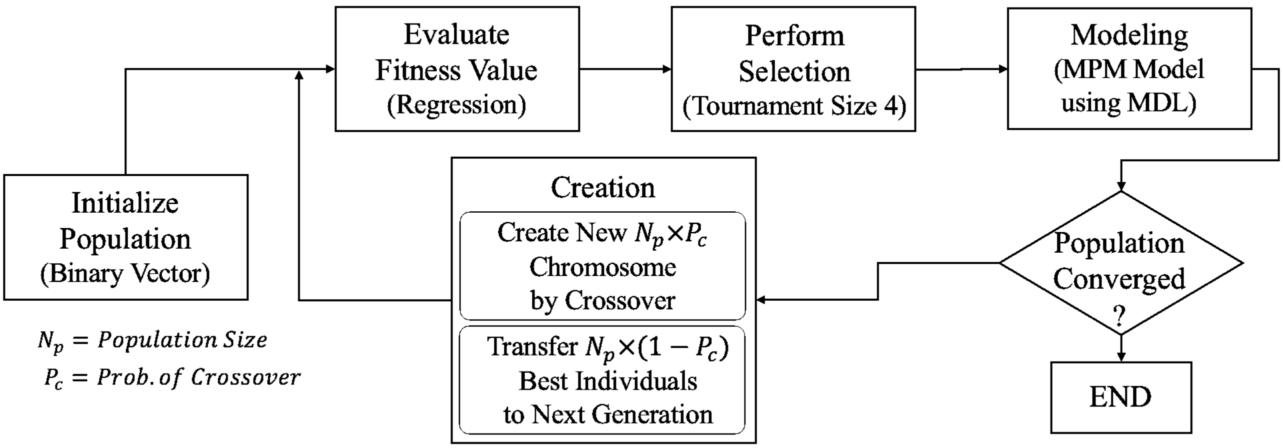


FIGURE 1. Procedure of proposed feature selection algorithm.

$Feature_1$	$Feature_2$	$Feature_3$		$Feature_{n-2}$	$Feature_{n-1}$	$Feature_n$
1	0	1	...	0	1	1

FIGURE 2. Example of a solution vector.

good distribution as an optimization problem that minimizes the probability model and population representation [15]. The probability model of ECGA is marginal product models (MPMs) developed as a product of marginal distributions on a partition of the genes [16]. Those models permit a direct linkage map with each partition separating tightly linked genes [15]. The proposed framework has been effectively solving interpretable relationships among features with evolutionary approach using minimum description length (MDL) model [17]. Next, Figure 1 the proposed algorithm is depicted in detail.

At the first step, the population is initialized as a candidate pool made up of the subset of features. Each individual of population utilizes a binary vector representation considered to obtain the relevant features [14]. Here, the length of its vector is set to the number of features in the given problem and each component of this vector is matched with each feature. And then, each component of this vector indicates that '1' means a particular feature in selected and '0' means a feature not selected in the subset. The example of the solution vector is depicted as Figure 2, where n is the total number of features in the given problem. The second stage of evaluation is to determine the ability of an individual to compete against others. It measures a fitness score appropriate to the combination of features to everyone which will be selected for reproduction in the next generation. The third phase of selection is to choose the fittest individuals and let them pass their genes to the next generation. Next, the creation of a new population based on MPM modeling using MDL. The definition of MPM is defined as a constrained optimization problem [17],

$$Minimize\ C_m + C_p \tag{1}$$

$$Subject\ to\ 2^{l_{bb,i}} \leq N_p \quad [1, N_{bb}] \tag{2}$$

where C_m is the model complexity representing the cost of a complex model and is given by

$$C_m = \log(N_p + 1) \sum_{i=1}^{N_{bb}} (2^{l_{bb,i}} - 1) \tag{3}$$

and C_p is the compressed population complexity representing the cost of using a simple as against a complex one and is evaluated as

$$C_p = \sum_{i=1}^{N_{bb}} \sum_{j=1}^{2^{l_{bb,i}}} N_{i,j} \log_2(\frac{N_p}{N_{i,j}}) \tag{4}$$

$N_b b$ is the number of BBs, $l_{bb,i}$ is the length of BB $i \in [1, N_{bb}]$ and $N_{i,j}$ is the number of individuals of the current population which possesses bit-sequence $j \in [1, 2^{l_{bb,i}}]$ for BB i [16]. The next population created by the optimal MPM in the following manner of population of size $N_p(1 - P_c)$ where P_c is the crossover probability is filled by the best individuals in the current population [17]. The rest $N_p P_c$ chromosomes are generated by randomly choosing subsets which are the gene groups identified by the current MPM. Thanks to the optimal MPM, we can achieve interpretable feature selection result in the given problem. Algorithm 1 describes an example of the proposed method, where N is the population size, S_i genes is represented the i^{th} subset, and P_k is the probability of observing outcome k. In Figure 3, we can find an optimal subset of features $[F_1, F_3][F_2][F_4]$ having the minimum value of combined complexity. In other words, we can interpret relationship among features that F_1 and F_3 are dependent relationship, and F_2 and F_4 don't have any relationship with the others.

Through the upper operation process, the proposed algorithm is able to find the optimal feature set and combination

TABLE 1. Simulation result of combined cycle power plant data set.

Feature Selection	Regression	mean	std	Q0	Q1	Q2	Q3	Q4	CI(Upper)	CI (Lower)
FS	Linear	20.793	0.9536	18.8064	20.2395	20.8234	21.4099	23.5104	21.5073	20.5287
	BR	20.793	0.9535	18.8069	20.2391	20.8235	21.4095	23.5107	21.0573	20.5287
	DT	20.6401	1.5823	17.1338	19.5804	20.7063	21.555	24.0383	21.0787	20.2015
	RF	11.141	0.9889	9.2004	10.4073	11.1237	11.7412	13.4999	11.4151	10.8669
	Xgboost	10.0088	0.9265	8.2461	9.4062	9.8891	10.4277	12.3804	10.2656	9.752
	GBM	15.1181	0.9455	12.9753	14.4425	14.9968	15.7972	17.9084	15.3802	14.8561
FS_{ECGA}	Linear	20.7919	0.9526	18.8064	20.2395	20.8234	21.4099	23.5104	21.0559	20.5278
	BR	20.7918	0.9525	18.8069	20.2391	20.8235	21.4095	23.5107	21.0559	20.5278
	DT	19.7936	1.3679	16.4735	19.0482	19.6401	20.6667	23.2194	20.1728	19.4145
	RF	11.0174	0.9804	9.0467	10.2392	10.9554	11.5514	13.2525	11.2892	10.7456
	Xgboost	9.9552	0.9408	8.2316	9.2559	9.8262	10.3677	12.3804	10.216	9.6944
	GBM	15.1136	0.9467	12.9532	14.437	14.984	15.7927	17.9084	15.376	14.8512
	LGB	11.315	1.0022	8.9463	10.6442	11.3212	11.8331	13.6089	11.5928	11.0372
FS_{GA}	Linear	20.7919	0.9526	18.8064	20.2395	20.8234	21.4099	23.5104	21.0559	20.5278
	BR	20.7918	0.9525	18.8069	20.2391	20.8235	21.4095	23.5107	21.0559	20.5278
	DT	19.3267	1.3063	16.1018	18.6604	19.2771	20.1797	22.8321	196887	18.9649
	RF	10.9224	0.9839	8.9185	10.1173	10.8809	11.5021	13.1572	11.1951	10.6496
	Xgboost	9.9552	0.9408	8.2316	9.2559	9.8262	10.3677	12.3804	10.216	9.6944
	GBM	15.1131	0.9469	12.9493	14.4319	14.9838	15.7912	17.9084	15.3756	14.8506
	LGB	11.315	1.0022	8.9463	10.6442	11.3212	11.8331	13.6089	11.5928	11.0372
FS_{BPSO}	Linear	20.7919	0.9526	18.8064	20.2395	20.8234	21.4099	23.5104	21.0559	20.5278
	BR	20.7918	0.9525	18.8069	20.2391	20.8235	21.4095	23.5107	21.0559	20.5278
	DT	19.6618	1.3444	16.4915	18.8962	19.5601	20.6123	23.1748	20.0344	19.2891
	RF	11.0044	0.9734	9.0385	10.1781	10.9749	11.5851	13.2491	11.2742	10.7346
	Xgboost	9.9552	0.9408	8.2316	9.2559	9.8262	10.3677	12.3804	10.216	9.6944
	GBM	15.1134	0.9469	12.9498	14.4329	14.984	15.7912	17.9084	15.3759	14.851
	LGB	11.315	1.0022	8.9463	10.6442	11.3212	11.8331	13.6089	11.5928	11.0372
FS_{BWOA}	Linear	20.7919	0.9526	18.8064	20.2395	20.8234	21.4099	23.5104	21.0559	20.5278
	BR	20.7918	0.9525	18.8069	20.2391	20.8235	21.4095	23.5107	21.0559	20.5278
	DT	19.3044	1.3011	16.1928	18.5685	19.1583	20.2636	22.3274	19.6651	18.9438
	RF	10.9255	0.976	8.9272	10.1429	10.8697	11.4938	13.1694	11.196	10.655
	Xgboost	9.9552	0.9408	8.2316	9.2559	9.8262	10.3677	12.3804	10.216	9.6944
	GBM	15.113	0.9469	12.9489	14.4324	14.9838	15.7913	17.9084	15.3755	14.8506
	LGB	11.315	1.0022	8.9463	10.6442	11.3212	11.8331	13.6089	11.5928	11.0372
FS_{BGWO}	Linear	20.7927	0.9535	18.8064	20.2395	20.8234	21.4099	23.5104	21.057	20.5284
	BR	20.7918	0.9525	18.8069	20.2391	20.8235	21.4095	23.5107	21.0559	20.5278
	DT	19.2649	1.3208	16.0897	18.4325	19.1813	20.162	22.0576	19.631	18.8988
	RF	10.9122	0.9764	8.943	10.1277	10.8575	11.4803	13.1606	11.1828	10.6415
	Xgboost	9.9622	0.9418	8.2416	9.2559	9.8262	10.3851	12.3804	10.2233	9.7012
	GBM	15.113	0.9469	12.9489	14.4326	14.9838	15.7912	17.9084	15.3755	14.8506
	LGB	11.315	1.0022	8.9463	10.6442	11.3212	11.8331	13.6089	11.5928	11.0372

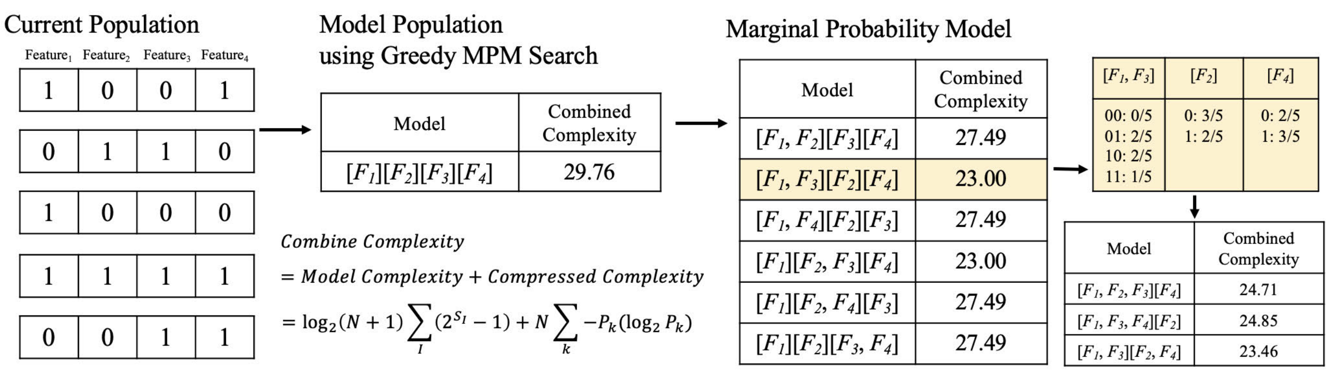


FIGURE 3. Example of MDL modeling.

which is the significant advantage of analyzing the relationship among all features.

IV. EXPERIMENTAL RESULTS

In this section, we compare and analyze the performance of the proposed evolutionary approach for interpretable feature selection method with four state-of-the-art metaheuristic-based feature selection methods in six types of real manufacturing data sets. For performance evaluation, the difference between the predicted value and the measured value is estimated using the root mean squared error (RMSE). Also, 80% of the total data is used as training data and the

TABLE 2. Simulation result of CPU data set.

Feature Selection	Regression	mean	std	Q0	Q1	Q2	Q3	Q4	CI(Upper)	CI (Lower)
FS	Linear	3300.2975	2360.982	929.3425	1613.9734	2463.2092	3478.1262	10444.6959	3954.7163	2645.8787
	BR	3202.5994	2225.8431	907.06	1504.5676	2538.9236	3502.3179	9688.0964	3819.5603	2585.6385
	DT	3694.9201	6383.8962	29.6667	527.1969	1236.4977	2699.2024	23116.4286	5464.4133	1925.4269
	RF	2598.6359	3611.0516	40.1542	395.8362	860.2705	1844.8109	14215.1816	3599.55	1597.7219
	XGBoost	1547.1167	1880.8893	39.3048	225.4697	973.9778	1534.7135	7735.6578	2068.43	1025.7704
	GBM	1876.4384	3226.4853	13.9602	99.7747	356.7714	1585.4508	15679.9601	2770.758	982.1187
	LGB	7356.0498	6205.2013	566.1045	2709.717	5649.0926	8655.214	28598.3917	9076.0122	5636.0873
FS_{ECGA}	Linear	2936.1442	1942.9215	784.858	1583.6207	2453.619	3429.6999	9639.6914	3474.6847	2397.6037
	BR	2875.3843	1907.0218	781.171	1498.0594	2288.3464	3395.5795	9684.1111	3403.9741	2346.7946
	DT	1151.6045	1057.7147	30.6429	352.1012	866.881	1720.4168	4650.7366	1444.7826	858.4263
	RF	1512.6249	2084.4842	36.8926	205.3209	572.4026	1445.8043	8176.5891	2090.4038	934.8459
	Xgboost	1106.2721	1325.7127	31.3828	137.3788	582.3327	1340.2565	5763.8137	1473.7341	738.81
	GBM	1250.8317	1748.2867	13.2882	60.4373	316.4216	1055.4372	6460.0342	1735.4231	766.2402
	LGB	7019.6749	6491.0882	292.9326	2242.4207	8621.1771	28625.1783	8818.87797	8818.8797	5220.4701
FS_{GA}	Linear	2514.008	1708.0489	676.3265	1348.6842	1952.4319	2967.3796	9636.3894	2987.4463	2040.5697
	BR	2713.4656	1872.9331	783.8125	1504.664	2048.5512	3140.818	9688.0964	3250.6066	2212.3246
	DT	1063.8126	1062.0699	13.0488	194.9802	756.2871	1603.0001	4664.6429	1358.1979	769.4272
	RF	1102.6728	1458.2061	27.5519	121.418	310.9734	1249.2488	5654.0944	1506.8595	698.4862
	Xgboost	871.9423	1052.5757	15.2531	107.8564	392.6294	1294.0659	4392.3563	1163.6961	580.1886
	GBM	718.2848	960.1977	11.1944	34.6672	238.534	798.129	3849.8669	984.4332	452.1365
	LGB	6762.2965	6549.7986	243.8169	1544.8292	5069.0225	8652.9639	28598.3917	8577.7747	4946.8183
FS_{BPSO}	Linear	2489.4365	1678.0226	747.9661	1360.0144	1950.1098	2931.5069	9639.6914	2954.552	2024.3209
	BR	2250.5511	1700.6772	666.797	1413.8204	1965.4594	3041.7771	9684.1111	3021.9461	2079.1561
	DT	946.3714	985.9802	30.0357	214.8675	501.7037	1293.6504	4549.4535	1219.6662	673.0767
	RF	1101.3957	1515.3769	28.766	134.7346	377.5921	1212.9358	7121.4542	1521.429	681.3624
	Xgboost	772.1282	922.0408	15.2531	96.1011	296.1754	957.6889	4077.5705	1027.7002	516.5563
	GBM	673.537	870.8013	12.0249	48.7198	218.8864	805.2182	3016.1698	914.9064	432.1676
	LGB	6650.6916	6593.0712	243.8169	1411.0345	4860.8647	8621.1771	28598.3917	8478.1641	4823.2191
FS_{BWOA}	Linear	2482.8057	1704.8166	676.3265	1348.6842	1924.0707	2931.5069	9636.3894	2955.3481	2010.2633
	BR	2544.6725	1722.9859	666.797	1415.961	1984.0909	3028.2939	9684.1111	3022.2511	2067.094
	DT	1013.5077	1023.7522	21.4695	171.7619	689.6353	1456.2649	4533.3333	1297.2721	729.7433
	RF	979.4969	1368.42	25.0149	110.9146	309.5575	1115.7322	6226.266	1358.7966	600.1972
	Xgboost	772.148	910.1911	15.2531	119.3069	382.2572	933.1891	4077.5705	1024.4355	519.8606
	GBM	670.2604	891.6103	11.3705	44.0339	206.4365	806.5223	3016.1698	917.3976	423.1232
	LGB	6724.8527	6605.5013	321.2631	1436.1385	4879.9901	8621.1771	28598.3917	8555.7706	4893.9348
FS_{BGWO}	Linear	2750.5234	1892.4037	676.3265	1378.1069	2017.3055	3504.6472	9636.3894	3275.0613	2225.9855
	BR	2720.6256	1840.8403	666.797	1460.745	2168.0731	3169.348	9684.1111	3230.8711	2210.38
	DT	1125.5646	1122.3833	26.8333	333.5995	746.1613	1804.7455	4689.9762	1436.6676	814.4615
	RF	1194.4604	1705.5125	23.7878	125.2744	366.4665	1102.8619	7196.0772	1672.1957	726.7252
	Xgboost	848.6559	1026.9995	26.2928	103.3777	321.0841	1159.411	4315.2159	1133.3205	563.9914
	GBM	762.0509	918.2088	10.9873	47.0511	333.0932	1156.2189	3016.1698	1016.5607	507.541
	LGB	6869.8863	6713.2755	250.7922	1710.4801	4865.9187	8655.214	31686.836	8730.6771	5009.0955

Algorithm 1 Minimum Description Length (MDL) Compute Model and Population Complexity

N_p = Size of population
L = List of building block lengths
C_m = List of building block lengths
LOOP Process
for i in number of building blocks **do**
  l_i = length of building block i
  for j in 2^{l_i} **do**
    $N_{i,j}$ = chromosomes in current population where $j \in [1, 2^{l_i}]$
    $C_p = \sum N_{i,j} \times \log_2(\frac{N_p}{N_{i,j}})$
  end for
end for
$MDL = C_m + C_p$

remaining 20% of the total data is used as test data to verify the performance of the proposal and referred feature selection methods. In addition, by repeating the experiment 50 times under the same conditions, we intend to secure statistical confidence in the results by utilizing the mean value, standard deviation, and quartile of the repeated experiment. At first, in two types of evolution-based algorithms named FS_{ECGA} and FS_{GA}, the size of a population and the generation are set to $pop\_size = 20 \times \log(Dim)$ and $gen = 1.5 \times \sqrt{pop\_size}$, respectively. In case of FS_{GA}, the crossover and mutation are fixed by 0.9 and 0.01. Next, in the binary particle swarm optimization-based feature selection of FS_{BPSO}, the number of particles is set to $10 \times pop\_size$. Lastly, the number of particles of binary whale optimization-based feature selection of FS_{BOWA} and binary gray wolf optimization-based feature selection of FS_{BGWO} are set to pop_size, equally. Moreover, to inspect the performance comparisons, we employ Random Forest (RF), Extreme Gradient Boosting (XGBoost), Gradient Boosting Machine (GBM) and Light-GBM that used default values of four supervised algorithms from the Scikit-learn library. We used PC equipped with an Intel(R) Core (TM) i7 6700, 340 Hz CPU and 32 GB RAM. All methods were experimented with codes written in Python.

TABLE 3. Simulation result of steel plates faults data set.

Feature Selection	Regression	mean	std	Q0	Q1	Q2	Q3	Q4	CI(Upper)	CI (Lower)
FS	Linear	9.7298	13.4029	3.0202	3.4268	3.6093	3.8976	51.3156	13.4448	6.0148
	BR	8.0587	10.0277	3.0111	3.4349	3.6138	3.8742	42.0031	10.8382	5.2793
	DT	4.2741	0.4952	3.3686	3.8918	4.1959	4.5758	5.5979	4.4113	4.1368
	RF	2.1217	0.1832	1.6661	1.9966	2.1374	2.2477	2.4579	2.1725	2.071
	Xgboost	2.2515	0.228	1.6383	2.074	2.2404	2.4473	2.6236	2.3147	2.1883
	GBM	2.3487	0.1853	1.9304	2.1972	2.3632	2.5008	2.7446	2.4001	2.2974
	LGB	2.1054	0.1928	1.6561	1.9705	2.0992	2.2579	2.5245	2.1588	2.0519
FS_{ECGA}	Linear	3.4829	0.1913	3.035	3.335	3.5034	3.6507	3.9342	3.5359	3.4299
	BR	3.509	0.1958	3.0054	3.3804	3.5117	3.6643	3.9631	3.5633	3.4548
	DT	3.5249	0.3113	2.6478	3.3368	3.567	3.7745	4.0902	3.611	3.4386
	RF	2.0029	0.1807	1.5595	1.8593	2.0092	2.1262	2.4036	2.053	1.9528
	Xgboost	2.1468	0.183	1.5317	2.0387	2.144	2.2691	2.4946	2.1975	2.096
	GBM	2.2575	0.1919	1.805	2.1343	2.2438	2.3977	2.7418	2.3107	2.2043
	LGB	2.0371	0.1979	1.516	1.8997	2.0303	2.2074	2.4206	2.0919	1.9822
FS_{GA}	Linear	3.4661	0.2033	2.9635	3.3191	3.4734	3.6033	3.8941	3.5225	3.4098
	BR	3.5045	0.2054	2.9882	3.3482	3.4954	3.6586	3.9012	3.5615	3.4476
	DT	3.1061	0.3665	2.3779	2.8312	3.0657	3.3145	4.0103	3.2077	3.0045
	RF	1.9049	0.1756	1.5767	1.7858	1.8966	2.033	2.3611	1.9536	1.8562
	Xgboost	1.9885	0.1958	1.4598	1.8432	1.9642	2.1365	2.4524	2.0428	1.9342
	GBM	2.1931	0.1815	1.7923	2.068	2.1893	2.3506	2.5898	2.2435	2.1428
	LGB	1.8934	0.1857	1.3498	1.7651	1.8667	2.0171	2.2601	1.9449	1.8419
FS_{BPSO}	Linear	3.4713	0.1957	3.0155	3.3125	3.4911	3.614	3.8945	3.5256	3.4171
	BR	3.4919	0.1968	2.9961	3.3606	3.4935	3.6494	3.8976	3.5464	3.4373
	DT	3.2414	0.3058	2.455	3.0354	3.2513	3.4871	4.0308	3.3262	3.1566
	RF	1.933	0.1802	1.5073	1.7838	1.9566	2.0536	2.2914	1.9829	1.883
	Xgboost	1.9846	0.1914	1.4757	1.8456	2.0115	2.119	2.3264	2.0377	1.9316
	GBM	2.2035	0.1855	1.7895	2.0557	2.2051	2.3525	2.5929	2.2549	2.152
	LGB	1.9211	0.1779	1.4748	1.7989	1.9099	2.0724	2.2388	1.9704	1.8718
FS_{BWOA}	Linear	3.4015	0.1946	2.9518	3.2392	3.4334	3.5572	3.8383	3.4555	3.3476
	BR	3.4303	0.1983	2.9608	3.2843	3.451	3.5974	3.8724	3.4852	3.3753
	DT	2.8631	0.367	1.9897	2.6798	2.9124	3.1269	3.5	2.9648	2.7614
	RF	1.794	0.1673	1.4004	1.6736	1.8287	1.9282	2.1481	1.8404	1.7477
	Xgboost	1.8544	0.2002	1.2903	1.697	1.8552	1.992	2.2898	1.9099	1.7989
	GBM	2.1269	0.1665	1.771	2.0087	2.1335	2.265	2.4927	2.1731	2.0808
	LGB	1.8015	0.1773	1.356	1.6686	1.8112	1.9432	2.1225	1.8506	1.7523
FS_{BGWO}	Linear	3.4162	0.1912	2.9764	3.2538	3.4563	3.564	3.8421	3.4692	3.3632
	BR	3.4562	0.1918	2.9681	3.3139	3.4755	3.5725	3.8825	3.5095	3.4032
	DT	2.9009	0.3829	2.09	2.6398	2.8893	3.1733	3.6632	3.0071	2.7948
	RF	1.8592	0.1818	1.4263	1.7311	1.8791	1.9683	2.3027	1.9096	1.8088
	Xgboost	1.9157	0.1982	1.4282	1.8145	1.9082	2.0586	2.3979	1.9706	1.8607
	GBM	2.161	0.184	1.7475	2.0396	2.1496	2.2919	2.5923	2.212	2.11
	LGB	1.8686	0.1697	1.3773	1.7501	1.8671	2.0141	2.2056	1.9156	1.8215

A. PREDICTIONS BY REGRESSION

In general, for predicting potential problems in manufacturing industries, seven well-known regressions are utilized as follows. Firstly, the stepwise linear regression is a method of linearly modeling the relationship among variables when the dependent variable is numerical. Secondly, Bayesian linear regression is an approach to statistical within the context of Bayesian inference. Particularly, the prior distribution is assumed then explicit results should be available for the posterior probability distributions of its parameters [4]. Thirdly, Decision tree builds regression models using a tree structure. It splits a dataset into groups of subsets while an associated decision tree is incrementally developed [8]. Fourthly, Random Forest (RF) is an algorithm that predicts or classifies based on the mode after making two or more decision trees. If only one decision tree is used, the probability of overfitting is high. Efficiently to solve this problem, one randomly constructs several trees, sees what results each have, and collects the results of each tree to predict the results. This technique has high consistency in analysis [4]. Lastly, Gradient boosted machines (GBM) refer to a class of ensemble machine learning algorithms constructed from decision tree models for classification or regression problems. GBM is suitable using a differentiable loss function and a gradient descent optimization algorithm in the direction of which trees are added one at a time and correct for the prediction error of the previous model [8].

B. EXPERIMENTAL RESULTS OF MANUFACTURING DATA SETS

Big data analysis, especially in the manufacturing industry, extracts meaningful values from large amounts of structured or unstructured data sets. For all data sets used in this study, performance was verified by converting unstructured data into structured data through various data preprocessing processes [18].

TABLE 4. Simulation result of steel industry energy consumption data set.

Feature Selection	Regression	mean	std	Q0	Q1	Q2	Q3	Q4	CI(Upper)	CI (Lower)
FS	Linear	0.2715	0.0036	0.2632	0.2691	0.2715	0.274	0.2794	0.2725	0.2705
	BR	0.2717	0.0035	0.2633	0.2693	0.2713	0.2743	0.2796	0.2727	0.2707
	DT	0.1912	0.0053	0.1153	0.1215	0.1255	0.1286	0.141	0.1269	0.1239
	RF	0.0705	0.0024	0.0658	0.0687	0.0706	0.0724	0.0754	0.0711	0.0698
	Xgboost	0.077	0.0021	0.072	0.0757	0.0771	0.0787	0.082	0.0776	0.0764
	GBM	0.0911	0.0022	0.0871	0.0893	0.0912	0.0927	0.0954	0.0917	0.0904
	LGB	0.0749	0.0023	0.0704	0.0731	0.075	0.0766	0.0796	0.0755	0.0742
FS_{ECGA}	Linear	0.2837	0.0034	0.2749	0.2815	0.2837	0.2855	0.2908	0.2846	0.2827
	BR	0.2837	0.0034	0.275	0.2815	0.2837	0.2855	0.2908	0.2846	0.2828
	DT	0.1152	0.0038	0.1061	0.01123	0.1152	0.1182	0.1225	0.1163	0.1141
	RF	0.0724	0.0025	0.0678	0.0707	0.0721	0.0748	0.0772	0.0731	0.0717
	Xgboost	0.0787	0.0023	0.0742	0.0768	0.0785	0.0807	0.0836	0.0793	0.078
	GBM	0.0917	0.0023	0.0877	0.0899	0.0916	0.0936	0.0957	0.0923	0.091
	LGB	0.076	0.0023	0.0718	0.0742	0.0763	0.0777	0.0808	0.7067	0.0754
FS_{GA}	Linear	0.2718	0.0035	0.2634	0.2695	0.2715	0.2734	0.2797	0.2728	0.2708
	BR	0.2726	0.0043	0.2653	0.2694	0.2718	0.2748	0.2887	0.2737	0.2714
	DT	0.1236	0.0062	0.1076	0.1198	0.1234	0.1281	0.1354	0.1253	0.1219
	RF	0.0701	0.0024	0.0652	0.0681	0.0703	0.0722	0.0747	0.0707	0.0694
	Xgboost	0.0766	0.0023	0.0721	0.0752	0.0767	0.0786	0.082	0.0773	0.076
	GBM	0.0908	0.0023	0.0866	0.089	0.0907	0.0925	0.0954	0.0914	0.0901
	LGB	0.0742	0.0021	0.0699	0.0724	0.0743	0.0756	0.079	0.0747	0.0736
FS_{BPSO}	Linear	0.2715	0.0036	0.2632	0.2691	0.2714	0.2741	0.2795	0.2725	0.2705
	BR	0.2717	0.0035	0.2635	0.2693	0.2713	0.2743	0.2797	0.2727	0.2707
	DT	0.1173	0.0053	0.1066	0.1141	0.1171	0.1211	0.1307	0.1187	0.1158
	RF	0.0702	0.0024	0.0659	0.0682	0.0707	0.0723	0.075	0.0709	0.0696
	Xgboost	0.0763	0.0022	0.0717	0.0748	0.0763	0.0782	0.081	0.0769	0.0757
	GBM	0.0906	0.0024	0.0866	0.0886	0.0905	0.0923	0.0951	0.0913	0.09
	LGB	0.0745	0.0022	0.0701	0.0728	0.0748	0.076	0.0795	0.0751	0.0739
FS_{BWOA}	Linear	0.2714	0.0036	0.2632	0.2691	0.2712	0.274	0.2794	0.2724	0.2705
	BR	0.2718	0.0043	0.2633	0.2693	0.2713	0.2742	0.29	0.273	0.2706
	DT	0.115	0.0053	0.1062	0.1104	0.1148	0.1184	0.1282	0.1165	0.1135
	RF	0.0697	0.0024	0.0651	0.0678	0.0698	0.0719	0.0747	0.0704	0.0691
	Xgboost	0.076	0.0022	0.0711	0.0747	0.0762	0.0777	0.081	0.0766	0.0754
	GBM	0.0906	0.0023	0.0866	0.0866	0.0906	0.0922	0.095	0.0912	0.0899
	LGB	0.0742	0.0021	0.0701	0.0724	0.0743	0.0756	0.079	0.0747	0.0736
FS_{BGWO}	Linear	0.2718	0.0039	0.2633	0.2692	0.2716	0.2742	0.2832	0.2729	0.2707
	BR	0.2721	0.0045	0.2633	0.2693	0.2715	0.2743	0.2919	0.2733	0.2708
	DT	0.1175	0.0056	0.1028	0.1129	0.1179	0.1215	0.1262	0.119	0.116
	RF	0.0697	0.0025	0.0649	0.0678	0.0697	0.0718	0.0755	0.0704	0.069
	Xgboost	0.0763	0.0023	0.0721	0.0748	0.0764	0.0778	0.8016	0.077	0.0757
	GBM	0.0907	0.0023	0.0866	0.0891	0.0906	0.0924	0.0951	0.0913	0.0901
	LGB	0.0743	0.0022	0.0704	0.0725	0.0744	0.0758	0.0794	0.0749	0.0737

1) COMBINED CYCLE POWER PLANT DATA SET

The first data consists of the hourly electrical energy output (PE) of the combined cycle power plant over the period 2006 - 2011, and the average ambient temperature (AT), ambient pressure (AP), relative humidity (RH), and evacuation vacuum (V) [19]. Table 1, the proposed evolutionary approach feature selection with XGBoost showed the best performance compared to meta-heuristic algorithms; that is, average is 9.9552 and 95% confidential interval is (9.6944, 10.216). On the other hand, the original feature selections without meta-heuristics showed lower performance than others. Here, each result of feature selection in meta-heuristic approaches is FS_{ECGA} = [AT,AP,RH,V], FS_{GA} = [AT,AP,RH,V], FS_{BWOA} = [AT,AP,RH], and FS_{BGWO} = [AT,AP,RH], respectively. Especially, the proposed interpretable feature selection algorithm provided optimal feature combination of [AT,AP][RH,V]. Thanks to those relationships, it could be possible to explain relations between features in the given problem.

2) CPU PERFORMANCE DATA SET

The data set of CPU performance is composed of total 6 integer input variables including machine cycle time (MYCT), minimum main memory (MMIN), maximum main memory (MMAX), cache memory (CACH), minimum channels in units (CHMIN), maximum channels in units (CHMAX) to predict the dependent variable, estimated relative performance (ERP) [20]. In Table 2, four metaheuristics based warping methods showed the best performance compared to the proposed evolutionary-based approach, while original methods showed low consistency. For an example, all of algorithms selected four features (i.e., MMIN, MMAX, CACH, CHMAX). Here, it could merely recognize whether features selected or not. To overcome its issue, the proposed feature selection algorithm found the optimal subset of features based on feature relationship model like [MYCT][MMAX][CACH][CHMIN][MMIN,CHMAX]. Here, while MMIN and CHMAX are dependent, the others are independent.

TABLE 5. Simulation result of productivity prediction of garment employees data set.

Feature Selection	Regression	mean	std	Q0	Q1	Q2	Q3	Q4	CI(Upper)	CI (Lower)
FS	Linear	0.0025	0.0027	0.0172	0.0208	0.0226	0.0241	0.0291	0.0232	0.0218
	BR	0.0025	0.0026	0.0173	0.0209	0.0227	0.024	0.0287	0.0232	0.0218
	DT	0.0279	0.0042	0.0192	0.0252	0.0279	0.0309	0.0385	0.0291	0.0268
	RF	0.0157	0.0027	0.0114	0.0134	0.0152	0.0182	0.0219	0.0164	0.0149
	Xgboost	0.017	0.0027	0.0119	0.0151	0.017	0.0192	0.0224	0.0178	0.0163
	GBM	0.0153	0.0023	1.0114	0.0135	0.0152	0.0168	0.0215	0.0159	0.0147
	LGB	0.0162	0.0026	0.0116	0.0142	0.016	0.0183	0.0221	0.017	0.0155
FS_{ECGA}	Linear	0.00229	0.0027	0.017	0.0207	0.0231	0.0247	0.0291	0.0236	0.0221
	BR	0.00229	0.0027	0.0171	0.0207	0.023	0.0247	0.0289	0.0236	0.0221
	DT	0.0197	0.0026	0.0152	0.0183	0.0192	0.0206	0.0263	0.0204	0.0189
	RF	0.0157	0.0022	0.0121	0.014	0.0154	0.0168	0.0214	0.0163	0.0151
	Xgboost	0.0164	0.0023	0.0115	0.0149	0.0164	0.0174	0.0214	0.0171	0.0158
	GBM	0.015	0.002	0.0117	0.0137	0.0149	0.0162	0.0199	0.0156	0.0144
	LGB	0.0155	0.0022	0.0121	0.0139	0.0149	0.0172	0.021	0.0161	0.0149
FS_{GA}	Linear	0.0225	0.0028	0.0167	0.0203	0.0226	0.0242	0.0294	0.0233	0.0218
	BR	0.0223	0.0025	0.0171	0.0205	0.0224	0.024	0.0278	0.023	0.0216
	DT	0.0182	0.0023	0.0146	0.0164	0.0178	0.0197	0.0229	0.0188	0.0176
	RF	0.0148	0.0022	0.0112	0.0132	0.0145	0.0163	0.019	0.0155	0.0142
	Xgboost	0.0158	0.0026	0.0107	0.0144	0.0161	0.0175	0.0205	0.0165	0.0151
	GBM	0.0147	0.0021	0.0114	0.0132	0.0147	0.0159	0.0201	0.0153	0.0142
	LGB	0.0151	0.0023	0.0107	0.0132	0.0149	0.0165	0.021	0.0157	0.0144
FS_{BPSO}	Linear	0.022	0.0026	0.0167	0.0202	0.0222	0.0241	0.0283	0.023	0.0215
	BR	0.0223	0.0026	0.0168	0.0203	0.0224	0.024	0.0282	0.023	0.0215
	DT	0.0178	0.0024	0.013	0.0162	0.0179	0.0193	0.0235	0.0185	0.0172
	RF	0.0148	0.0023	0.0107	0.0132	0.0149	0.0161	0.0198	0.0155	0.0142
	Xgboost	0.0156	0.0025	0.0109	0.0138	0.0158	0.0173	0.0209	0.0163	0.015
	GBM	0.0147	0.002	0.0116	0.0133	0.0146	0.0159	0.0201	0.0153	0.0142
	LGB	0.0149	0.0022	0.0108	0.0131	0.0148	0.0163	0.0198	0.0155	0.0143
FS_{BWOA}	Linear	0.0221	0.0026	0.0166	0.0202	0.0222	0.0239	0.028	0.0228	0.0214
	BR	0.0221	0.0026	0.0167	0.0203	0.0223	0.0238	0.0278	0.0228	0.0214
	DT	0.0171	0.0023	0.0126	0.0151	0.0173	0.0188	0.0216	0.0177	0.0165
	RF	0.0143	0.0023	0.0101	0.0128	0.0142	0.0157	0.0185	0.0149	0.0136
	Xgboost	0.015	0.0024	0.01	0.0136	0.015	0.0166	0.0191	0.0157	0.0143
	GBM	0.0143	0.002	0.0112	0.0128	0.0142	0.0156	0.0195	0.0149	0.0138
	LGB	0.0146	0.0021	0.0107	0.0131	0.0144	0.016	0.0198	0.0152	0.014
FS_{BGWO}	Linear	0.0222	0.0027	0.0167	0.0202	0.0224	0.0239	0.0289	0.0229	0.0214
	BR	0.0222	0.0026	0.0169	0.0204	0.0224	0.0239	0.0281	0.0229	0.0215
	DT	0.0177	0.0025	0.0125	0.0155	0.0175	0.0196	0.0232	0.0184	0.017
	RF	0.0144	0.0024	0.0104	0.0128	0.0141	0.016	0.0194	0.015	0.0137
	Xgboost	0.0155	0.0025	0.0105	0.0138	0.0153	0.0171	0.02	0.0161	0.0148
	GBM	0.0145	0.0021	0.0112	0.0132	0.0143	0.0157	0.0199	0.0151	0.0139
	LGB	0.0151	0.0024	0.0107	0.0135	0.0146	0.0166	0.0218	0.0157	0.0144

3) STEEL PLATES FAULTS DATA SET

The data set divides stainless steel sheet surface defects into 7 types such as Pastry, Z_Scratch, K_Scatch, Stains, Dirtiness, Bumps, and Other_Faults [21]. It consists of 27 features describing their geometric shapes and contours such as Min_X (1), Max_X (2), Min_Y (3), Max_Y (4), Pixels_Areas (5), Perimeter_X (6), Perimeter_Y (7), Sum_Luminosity (8), Minimum_Luminosity (9), Maximum_Luminosity (10), Length_Conveyer (11), Steel_A300 (12), Steel_A400 (13), Steel_Plate_Thickness (14), Edges_Index (15), Empty_Index (16), Square_Index (17), Outside_X_Index (18), Edges_X_Index (19), Edges_Y_Index (20), Outside_Global_Index (21), LogOfAreas (22), Log_X_Index (23), Log_Y_Index (24), Orientation_Index (25), and Luminosity_Index (26). Type. In Table 3, two meta heuristic based warping methods (i.e., FS_{BWOA}, FS_{BGWO}) showed little better performances compared to the proposed feature selection algorithm. For one example, all meta-heuristic approaches could find each subset of features. Firstly, FS_{ECGA} selected 14 features such as [3, 5, 8, 10, 11, 12, 13, 14, 15, 16, 19, 20, 21, 25]. Secondly, FS_{GA} selected 11 features such as [1, 4, 9, 13, 14, 16, 21, 22, 24, 25, 26]. Thirdly, FS_{BSPO} selected 14 features such as [2, 3, 4, 5, 6, 8, 10, 11, 12, 15, 17, 18, 19, 23]. Fourthly, FS_{BWOA} selected 17 features such as [1, 2, 4, 5, 8, 9, 11, 13, 14, 15, 16, 19, 21, 22, 24, 25]. Finally, FS_{BGWO} selected 18 features such as [1, 3, 9, 11, 12, 13, 14, 15, 16, 20, 21, 22, 23, 24, 25, 26]. Here, the proposed algorithm could achieve the optimal solution based on four kinds of feature relationship such as [Y_Maximum(4), SteelA300(12)], [X_Maximum(2), Outside_X_Index(18)], [Steel Plate Thickness(14), Log X index(23)], [Minimum Luminosity(9), Orientation Index(25)]. From those relations, it effectively described the relationship among features.

4) STEEL INDUSTRY ENERGY CONSUMPTION DATA SET

This data provided by Daewoo Steel in Gwangyang of South Korea produces several types of coils, steel plates

TABLE 6. Simulation result of real manufacturing process data set.

Feature Selection	Regression	mean	std	Q0	Q1	Q2	Q3	Q4	CI(Upper)	CI (Lower)
	Linear	1.4579	0.2006	0.9953	1.3052	1.4802	1.6334	1.9071	1.5135	1.4023
	BR	1.435	0.197	0.9829	1.2745	1.4618	1.6112	1.7602	1.4896	1.3804
FS	DT	2.1522	0.3668	1.5434	1.9031	2.0806	2.3569	3.6226	2.2539	2.0505
	RF	1.1294	0.2241	0.7775	0.9555	1.0972	1.286	1.7284	1.1915	1.0672
	Xgboost	1.3042	0.257	0.8709	1.1201	1.2719	1.4504	1.9494	1.3754	1.233
	GBM	1.2378	0.2516	0.7813	1.0701	1.1925	1.3929	1.8921	1.3076	1.1681
	LGB	0.0162	0.0026	0.886	1.0917	1.224	1.3295	1.8068	1.2716	1.1695
	Linear	1.3777	0.1835	0.9637	1.2329	1.4034	1.5266	1.6982	1.4285	1.3268
	BR	1.3747	0.1831	0.9632	1.2296	1.4025	1.5284	1.6849	1.4254	1.3239
FS_{ECGA}	DT	1.6139	0.1811	1.2258	1.4893	1.6447	1.7303	2.012	1.6641	1.5637
	RF	1.0172	0.1688	0.7156	0.8924	1.0083	1.1375	1.3751	1.064	0.9704
	Xgboost	1.1101	0.1735	0.7494	0.9812	1.0784	1.2556	1.4244	1.1582	1.062
	GBM	1.0569	0.1632	0.7785	0.9242	1.0617	1.1923	1.419	1.1021	1.0116
	LGB	1.1046	0.1679	0.7305	0.9724	1.111	1.2278	1.4648	1.1511	1.0581
	Linear	1.341	0.1799	0.9492	1.1975	1.3728	1.4713	1.643	1.3909	1.2911
	BR	1.3747	0.1831	0.9632	1.2296	1.4025	1.5284	1.6849	1.4254	1.3239
FS_{GA}	DT	1.6139	0.1811	1.2258	1.4893	1.6447	1.7303	2.012	1.6641	1.5637
	RF	0.9731	0.1628	0.678	0.8568	0.9739	1.1139	1.3134	1.0185	0.9277
	Xgboost	1.0199	0.1455	0.7043	0.9152	1.0084	1.1136	1.3485	1.0602	0.9796
	GBM	1.0086	0.1656	0.7185	0.8714	1.0039	1.1317	1.3834	1.0545	0.9627
	LGB	1.0503	0.1573	0.7674	0.9266	1.0632	1.1764	1.3209	1.0939	1.0067
	Linear	1.3412	0.1755	0.9562	1.1954	1.3742	1.4729	1.6607	1.3898	1.2925
	BR	1.3376	0.1764	0.9436	1.2056	1.3816	1.4731	1.6435	1.3865	1.2887
FS_{BPSO}	DT	1.4937	0.1819	1.1529	1.3505	1.4796	1.5936	1.9939	1.5441	1.4433
	RF	0.9785	0.1623	0.7307	0.8521	0.9534	1.108	1.2992	1.0235	0.9335
	Xgboost	1.0502	0.1628	0.761	0.9934	1.0435	1.1781	1.3736	1.0953	1.005
	GBM	1.0109	0.1639	0.7047	0.8774	1.0098	1.1328	1.3157	1.0563	0.9655
	LGB	1.0443	0.144	0.7659	0.922	1.0672	1.1673	1.2854	1.0843	1.0044
	Linear	1.2906	0.1788	0.9039	1.1596	1.3216	1.4206	1.6112	1.3401	1.241
	BR	1.2958	0.1799	0.909	1.1606	1.3257	1.4349	1.6111	1.3457	1.2459
FS_{BWOA}	DT	1.2871	0.2015	0.8684	1.1613	1.253	1.3904	1.7508	1.343	1.2312
	RF	0.935	0.1691	0.6497	0.7975	0.932	1.0663	1.3078	0.9819	0.8881
	Xgboost	0.9557	0.1428	0.6831	0.8562	0.9318	1.0627	1.2452	0.9953	0.9161
	GBM	0.9199	0.1506	0.635	0.8267	0.9162	1.0085	1.2495	0.9616	0.8781
	LGB	0.9448	0.1384	0.681	0.856	0.9348	1.0385	1.2407	0.9831	0.9064
	Linear	1.1395	0.1813	0.9322	1.1866	1.3513	1.4612	1.6392	1.3697	1.2692
	BR	1.3293	0.1826	0.9139	1.19	1.3538	1.4648	1.6439	1.3799	1.2787
FS_{BGWO}	DT	1.2844	0.1828	0.9647	1.1366	1.2862	1.4029	1.8226	1.3351	1.2338
	RF	0.9467	0.1531	0.6897	0.8229	0.9417	1.0526	1.2305	0.9892	0.9043
	Xgboost	1.003	0.1607	0.696	0.8722	0.9744	1.1126	1.4079	1.0475	0.9584
	GBM	0.9842	0.1581	0.7043	0.8875	0.9571	1.1044	1.303	1.0281	0.9404
	LGB	1.0036	0.1517	0.6926	0.8938	1.017	1.1142	1.2876	1.0457	0.9616

and steel plates [22]. Energy usage information by industry was obtained by collecting daily, monthly, and yearly data. There are each data feature as follows : Industry Energy Consumption Continuous(kWh), Lagging Current reactive power Continuous(kVarh), Leading Current reactive power Continuous (kVarh), tCO2(CO2) Continuous(ppm), Lagging Current power factor Continuous(%), Leading Current Power factor Continuous(%), Number of Seconds from midnight Continuous(S), Week status Categorical(Weekend or Weekday), Day of week Categorical(Sunday to Saturday), Load Type Categorical Light Load, Medium Load, Maximum Load [21]. Table 4, the proposed evolutionary approach feature selection with RF showed the best performance compared to meta-heuristic algorithms. For one example, all of algorithms selected most of features excluding one feature (i.e., Lagging Current reactive power Continuous). Here, the proposed approach found one dependent relation of Leading Current reactive power Continuous and Week status Categorical.

5) PRODUCTIVITY PREDICTION OF GARMENT EMPLOYEES DATA SET

The data relates to the apparel industry, a very labor-intensive industry. For this industry, the production and delivery performance of garment manufacturing company employees is critical to meeting global demand [23]. The data used in this study for regression purposes to predict the productivity range (0-1). The factors of this data are as follows: Date (1), Day of the Week (2), Quarter (3), Associated department with the instance (4), Associated team number with the instance (5), Number of workers in each team (6), Number of changes in the style of a particular product (7), Targeted productivity set by the Authority for each team for each day (8), Standard Minute Value (9), Work in progress (10) including the number of unfinished items for products, Represents the amount of overtime by each team in minutes (11), Represents the amount of financial incentive (in BDT) that enables or motivates a particular course of action (12), The amount of time when the production was interrupted due to several reasons (13), The number of workers who were idle due to production interruption l(14). The actual % of productivity that was delivered by the workers. It ranges from 0-1 [23]. Table 5, the proposed evolutionary approach feature selection with machine learning based regressions (RF, XBGoost, GBM, LGB) showed the best performance compared to meta-heuristic algorithms. For one example, all meta-heuristic approaches could find each subset of features. Firstly, FS_{GA}, FS_{BPSO}, FS_{BWOA}, and FS_{BGWO} selected 8 features whose position number was [1, 4, 5, 6, 7, 10, 11, 12]. However, the proposed algorithm of FS_{ECGA} selected 8 features whose position number was [2, 5, 6, 7, 8, 9, 10, 11]. Here, the proposed algorithm could achieve the optimal solution based on three dependent feature relationships such as [2, 5][6][] [8, 9][10, 11]. Based on those relations, it effectively describes the relationship among optimal features.

6) REAL MANUFACTURING PROCESS DATA SET

In general, data collected and utilized in the manufacturing process consists of various inputs. Thus, it is difficult to intuitively analyze the relationship between variables. Effectively to solve this problem, a preliminary analysis is performed on collected data based on the domain knowledge of the manufacturing process, but it is still almost impossible to analyze the relationship between factors. Also, classification and prediction analysis were not so different that the results were barely available for interpretation on the relationship between factors. The main difficulty is that it is implausible to derive an accurate result through classification or prediction based on the performed analysis. This section attempts to verify the performance of the algorithm by using 20 process factors and quality result data collected from a specific manufacturing company. Detailed explanations on the data, however, are excluded due to security issues in the manufacturing process of the company that provided the data. Table 6, the proposed evolutionary approach feature selection with machine learning based regressions (RF, XGBoost, GBM, LGB) showed the best performance compared to meta-heuristic algorithms. Firstly, FS_{ECGA}, FS_{GA}, FS_{BPSO}, and FS_{BGWO} selected 9 features whose position number was [4, 5, 7, 8, 9, 10, 15, 16, 17, 18, 19]. The other of FS_{BWOA} selected 11 features whose position number was [1, 2, 4, 7, 8, 9, 12, 14, 16, 19, 20]. Here, the proposed algorithm could achieve the optimal solution based on three kinds of feature relationship such as [3, 16][11, 12][18, 19]. From those relations, it effectively described the relationship among features.

V. CONCLUSION

In this paper, we propose a novel feature selection framework employing evolutionary approaches, especially estimation of distribution algorithm (EDA). Most of data analysis techniques affiliated with artificial intelligence domain includes characteristics of deriving features and patterns among numerous amounts of data. Meanwhile, reasoning out causal substance from astronomical number of variables and parameters in artificial intelligence models is significantly challenging. Accordingly, recent trend in artificial intelligence stresses about the importance of explainability regarding data training results. With the combined use of feature selection as well as evolutionary algorithm, we deliver logical outcome from the data analysis results through employing extended compact genetic algorithm (ECGA). By expanding the set of features candidate pool, we search for the most optimal subset of features combination using evolutionary algorithms. Particularly, ECGA is specialized in providing how such solution is derived based on probabilistic models during generation cycle. In other words, model blocks enable tracing back the cluster combination outcomes and interpreting the data analysis results as well. Our proposed approach shows meaningful performance compared to state-of-the-art metaheuristic based feature selection algorithms. We reckon that our proposed research has advantages of not only prominent performance in predictions but also logical background for data analysis result. In real manufacturing industries, the data set contains numerous features. In the feature selection problems, the scalability of the proposed algorithm for processing large data is important. Another important point is stability for solving feature selection problems because the proposed algorithm finds the same subset of features for different dataset samples. However, most feature selection algorithms become unstable due to the iterative process to find the optimal classification. Therefore, stability is just as important as classification accuracy.

REFERENCES

[1] P. Agrawal, H. F. Abutarboush, T. Ganesh, and A. W. Mohamed, "Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019)," *IEEE Access*, vol. 9, pp. 26766–26791, 2021.

[2] G. Dhiman, "ESA: A hybrid bio-inspired metaheuristic optimization approach for engineering problems," *Eng. Comput.*, vol. 37, no. 1, pp. 323–353, 2021.

[3] C. P. Ezenkwu, U. I. Akpan, and B. U.-A. Stephen, "A class-specific metaheuristic technique for explainable relevant feature selection," *Mach. Learn. Appl.*, vol. 6, Dec. 2021, Art. no. 100142.

[4] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, "Machine learning in manufacturing: Advantages, challenges, and applications," *Prod. Manuf. Res.*, vol. 4, no. 1, pp. 23–45, Jan. 2016.

[5] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, Jan. 2016.

[6] B. Iñigo, N. Colinas-Armijo, L. N. L. De Lacalle, and G. Aguirre, "Digital twin-based analysis of volumetric error mapping procedures," *Precis. Eng.*, vol. 72, pp. 823–836, Nov. 2021.

[7] M. A. Tawhid and A. M. Ibrahim, "Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 3, pp. 573–602, Mar. 2020.

[8] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *Social Netw. Comput. Sci.*, vol. 2, no. 3, p. 160, May 2021.

[9] A. Deniz, M. Angin, and P. Angin, "Evolutionary multiobjective feature selection for sentiment analysis," *IEEE Access*, vol. 9, pp. 142982–142996, 2021.

[10] W. Samek and K.-R. Müller, "Towards explainable artificial intelligence," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K.-R. Müller, Eds. Cham, Switzerland: Springer, 2019, pp. 5–22.

[11] C. W. Ahn, *Advances in Evolutionary Algorithms: Theory, Design and Practice*. Berlin, Germany: Springer, 2006.

[12] Y. Jin, H. Wang, T. Chugh, D. Guo, and K. Miettinen, ''Data-driven evolutionary optimization: An overview and case studies,'' *IEEE Trans. Evol. Comput.*, vol. 23, no. 3, pp. 442–458, Jun. 2019.

[13] J. Kennedy and R. Eberhart, ''Particle swarm optimization,'' in *Proc. Int. Conf. Neural Netw. (ICNN)*, vol. 4, Aug. 1995, pp. 1942–1948.

[14] S. Lee, S. Soak, S. Oh, W. Pedrycz, and M. Jeon, ''Modified binary particle swarm optimization,'' *Prog. Natural Sci.*, vol. 18, no. 9, pp. 1161–1166, 2008.

[15] P.-C. Hung and Y.-P. Chen, ''IECGA: Integer extended compact genetic algorithm,'' in *Proc. 8th Annu. Conf. Genetic Evol. Comput.*, New York, NY, USA, Jul. 2006, pp. 352–359.

[16] R. G. Harik, G. F. Lobo, and K. Sastry, ''Linkage learning via probabilistic modeling in the extended compact genetic algorithm (ECGA),'' in *Scalable Optimization via Probabilistic Modeling*, M. Pelikan, K. Sastry, E. CantúPaz, Eds. Berlin, Germany: Springer, 2006, pp. 39–61.

[17] M. Hauschild and M. Pelikan, ''An introduction and survey of estimation of distribution algorithms,'' *Swarm Evol. Comput.*, vol. 1, no. 3, pp. 111–128, 2011.

[18] D. Dua and C. Graff. (2019). *UCI Machine Learning Repository*. [Online]. Available: https://doi.org/10.6084/m9.figshare.853801

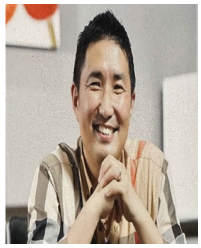
[19] H. Kaya, P. Tüfekci, and F. S. Gürgen, ''Local and global learning methods for predicting power of a combined gas & steam turbine,'' in *Proc. Int. Conf. Emerg. Trends Comput. Electron. Eng. (ICETCEE)*, 2012, pp. 13–18.

[20] P. Ein-Dor and J. Feldmesser. (1987). *Relative CPU Performance Data*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Computer+Hardware

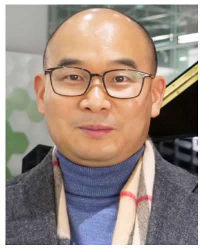
[21] *Dataset Provided by Semeion*, Research Center of Sciences of Communication, Via Sersale, Rome, Italy, 2010. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/steel+plates+faults

[22] V. Sathishkumar, J. Lim, M. Lee, K. Cho, J. Park, C. Shin, and Y. Cho, ''Industry energy consumption prediction using data mining techniques,'' *Int. J. Energy, Inf. Commun.*, vol. 11, no. 1, pp. 7–14, 2020.

[23] A. A. Imran, M. S. Rahim, and T. Ahmed, ''Mining the productivity data of the garment industry,'' *Int. J. Energy, Inf. Commun.*, vol. 19, no. 3, pp. 1743–8195, 2021.



SANGHOUN OH received the Ph.D. degree from the Department of Information and Communications, Gwangju Institute of Science and Technology (GIST), Republic of Korea, in 2011. From 2011 to 2018, he was with the Manufacturing Technology Center of Samsung Electronics, South Korea. He is currently the Head of Development/Package and Test/Management Data Analytic with the Department of Digital Transformation, SK Hynix. His research interests include genetic algorithms/programming, multi-objective optimization, and explainable artificial intelligence. He also concentrates on applying meta-heuristic algorithms into manufacturing industries to find optimized solutions.



CHANG WOOK AHN (Member, IEEE) received the Ph.D. degree from the Department of Information and Communications, Gwangju Institute of Science and Technology (GIST), Republic of Korea, in 2005. From 2005 to 2007, he was with the Samsung Advanced Institute of Technology, South Korea. From 2007 to 2008, he was a Research Professor with GIST. From 2008 to 2016, he was an Assistant/Associate Professor with the Department of Computer Engineering, Sungkyunkwan University (SKKU), Republic of Korea. He is currently a Professor with the School of Artificial Intelligence, GIST. His research interests include genetic algorithms/programming, multi-objective optimization, neural networks, and quantum machine learning.