APEM journal

Advances in Production Engineering & Management
Volume 20 | Number 4 | December 2025 | pp 458–474
https://doi.org/10.14743/apem2025.4.552

ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

A multi-objective feature selection and self-paced ensemble framework for semiconductor defect detection

Zheng, H.[a], Gao, X.[a,*], Yang, X.[a], Jing, G.[a], Yang, M.[a], Liu, Y.[a]

[a]School of Mechanical Engineering, Xi'an University of Technology, Xi'an, P.R. China

ABSTRACT

In semiconductor manufacturing, defect detection is commonly performed using high-dimensional process data. These data often exhibit class imbalance and class overlap, which create challenges for achieving reliable classification performance. To address these issues, this study proposes a multi-objective feature selection and self-paced ensemble (MOFS-SPE) framework. The framework employs a multi-objective evolutionary algorithm based on decomposition (MOEA/D) for feature selection. In this process, the area under the precision–recall curve (AUPRC) and the *R*-value are used as objective functions to identify feature subsets that are highly relevant to quality outcomes. In addition, the framework integrates the self-paced ensemble (SPE) with tree-based classifiers to handle imbalanced and overlapping data. Experiments conducted on a real semiconductor manufacturing dataset (SECOM dataset) demonstrate the effectiveness of the proposed approach. Compared with using the full feature set, the selected features increase the area under the receiver operating characteristic curve (AUROC) from 0.685 to 0.770 and the AUPRC from 0.932 to 0.972. When applying the SPE framework, the specificity of the decision tree model improves from 0.048 to 0.667, thereby enhancing the reliability of identifying defective products. Overall, the proposed framework provides a useful reference for intelligent quality inspection in semiconductor production environments.

ARTICLE INFO

Keywords:
Semiconductor manufacturing;
Defect detection;
Quality inspection;
Class imbalance;
Class overlap;
Multi-objective feature selection;
Self-paced ensemble;
Machine learning

**Corresponding author:*
xaut66gxq@163.com
(Gao, X.)

Article history:
Received 27 August 2025
Revised 30 November 2025
Accepted 3 December 2025

References

[1] Song, W.T., Huo, L. (2024). Machine learning for enhancing manufacturing quality control in ultrasonic nondestructive testing: A Wavelet Neural Network and Genetic Algorithm approach, *Advances in Production Engineering & Management*, Vol. 19, No. 3, 347-357, doi: 10.14743/apem2024.3.511.

[2] Nuhu, A.A., Zeeshan, Q., Safaei, B., Shahzad, M.A. (2023). Machine learning-based techniques for fault diagnosis in the semiconductor manufacturing process: A comparative study, *The Journal of Supercomputing*, Vol. 79, No. 2, 2031-2081, doi: 10.1007/s11227-022-04730-x.

[3] Kang, S., Kim, E., Shim, J., Chang, W., Cho, S. (2018). Product failure prediction with missing data, *International Journal of Production Research*, Vol. 56, No. 14, 4849-4859, doi: 10.1080/00207543.2017.1407883.

[4] He, Z., Hu, H., Zhang, M., Yin, X. (2022). Product quality prediction based on process data: Taking the gearbox of Company D as an example, *Quality Engineering*, Vol. 34, No. 3, 409-422, doi: 10.1080/08982112.2022.2063743.

[5] Carvajal Soto, J.A., Tavakolizadeh, F., Gyulai, D. (2019). An online machine learning framework for early detection of product failures in an Industry 4.0 context, *International Journal of Computer Integrated Manufacturing*, Vol. 32, No. 4-5, 452-465, doi: 10.1080/0951192X.2019.1571238.

[6] Jeong, E.-Y., Kim, J., Jang, W.-H., Lim, H.-C., Noh, H., Choi, J.-M. (2021). A more reliable defect detection and performance improvement method for panel inspection based on artificial intelligence, *Journal of Information Display*, Vol. 22, No. 3, 127-136, doi: 10.1080/15980316.2021.1876174.

[7] Kang, Z., Catal, C., Tekinerdogan, B. (2022). Product failure detection for production lines using a data-driven model, *Expert Systems with Applications*, Vol. 202, Article No. 117398, doi: 10.1016/j.eswa.2022.117398.

[8] Wang, N., Li, P., Zhao, Z., Yang, J. (2022). Quality abnormal recognition model based on convolutional neural network, *Operations Research and Management Science*, Vol. 31, 220-225, doi: 10.12005/orms.2022.0205.

[9] Hong, Y., Yang, M., Jiang, Y., Du, D., Chang, B. (2023). Real-time quality monitoring of ultrathin sheets edge welding based on microvision sensing and SOCIFS-SVM, *IEEE Transactions on Industrial Informatics*, Vol. 19, No. 4, 5506-5516, doi: 10.1109/TII.2022.3199258.

[10] Ismail, M., Mostafa, N.A., El-assal, A. (2022). Quality monitoring in multistage manufacturing systems by using machine learning techniques, *Journal of Intelligent Manufacturing,* Vol. 33, No. 8, 2471-2486, doi: 10.1007/s10845-021-01792-1.

[11] Santos, M.S., Abreu, P.H., Japkowicz, N., Fernández, A., Santos, J. (2023). A unifying view of class overlap and imbalance: Key concepts, multi-view panorama, and open avenues for research, *Information Fusion*, Vol. 89, 228-253, doi: 10.1016/j.inffus.2022.08.017.

[12] Hancock, J.T., Khoshgoftaar, T.M., Johnson, J.M. (2023). Evaluating classifier performance with highly imbalanced Big Data, *Journal of Big Data*, Vol. 10, No. 1, Article No. 42, doi: 10.1186/s40537-023-00724-5.

[13] Borsos, Z., Lemnaru, C., Potolea, R. (2018). Dealing with overlap and imbalance: A new metric and approach, *Pattern Analysis and Applications*, Vol. 21, No. 2, 381-395, doi: 10.1007/s10044-016-0583-6.

[14] Liu, Z., Cao, W., Gao, Z., Bian, J., Chen, H., Chang, Y., Liu, T.Y. (2020). Self-paced ensemble for highly imbalanced massive data classification, In: *Proceedings of 2020 IEEE 36th International Conference on Data Engineering (ICDE),* Dallas, USA, 841-852, doi: 10.1109/ICDE48307.2020.00078.

[15] Zhou, C., Fang, X. (2023). A convex two-dimensional variable selection method for the root-cause diagnostics of product defects, *Reliability Engineering & System Safety*, Vol. 229, Article No. 108827, doi: 10.1016/j.ress.2022.108827.

[16] Sun, C., Yin, Y., Kang, H., Ma, H. (2022). A distributed principal component regression method for quality-related fault detection and diagnosis, *Information Sciences*, Vol. 600, 301-322, doi: 10.1016/j.ins.2022.03.069.

[17] Liu, K., Hu, X., Zhou, H., Tong, L., Widanage, W.D., Marco, J. (2021). Feature analyses and modeling of lithium-ion battery manufacturing based on random forest classification, *IEEE/ASME Transactions on Mechatronics*, Vol. 26, No. 6, 2944-2955, doi: 10.1109/TMECH.2020.3049046.

[18] Saqlain, M., Jargalsaikhan, B., Lee, J.Y. (2019). A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 32, No. 2, 171-182, doi: 10.1109/TSM.2019.2904306.

[19] Mlinarič, J., Pregelj, B., Boškoski, P., Dolanc, G., Petrovčič, J. (2024). Optimization of reliability and speed of the end-of-line quality inspection of electric motors using machine learning, *Advances in Production Engineering & Management,* Vol. 19, No. 2, 182-196, doi: 10.14743/apem2024.2.500.

[20] Azamfar, M., Li, X., Lee, J. (2020). Deep learning-based domain adaptation method for fault diagnosis in semiconductor manufacturing, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 33, No. 3, 445-453, doi: 10.1109/TSM.2020.2995548.

[21] Stavropoulos, P., Papacharalampopoulos, A., Stavridis, J., Sampatakakis, K. (2020). A three-stage quality diagnosis platform for laser-based manufacturing processes, *The International Journal of Advanced Manufacturing Technology*, Vol. 110, No. 11, 2991-3003, doi: 10.1007/s00170-020-05981-9.

[22] Chen, Z., Maske, H., Shui, H., Upadhyay, D., Hopka, M., Cohen, J., Lai, X., Huan, X., Ni, J. (2023). Stochastic deep Koopman model for quality propagation analysis in multistage manufacturing system, *Journal of Manufacturing Systems*, Vol. 71, 609-619, doi: 10.1016/j.jmsy.2023.10.012.

[23] Chawla, N.V., Lazarevic, A., Hall, L.O., Bowyer, K.W. (2003). SMOTEBoost: Improving prediction of the minority class in boosting, In: *Proceedings of PKDD 2003 – 7th European Conference on Principles and Practice of Knowledge Discovery in Databases,* Cavtat-Dubrovnik, Croatia, 107-119, doi: 10.1007/978-3-540-39804-2_12.

[24] Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A. (2009). RUSBoost: A hybrid approach to alleviating class imbalance, *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* Vol. 40, No. 1, 185-197, doi: 10.1109/TSMCA.2009.2029559.

[25] Wang, S., Yao, X. (2009). Diversity analysis on imbalanced data sets by using ensemble models. In: *Proceedings of 2009 IEEE Symposium on Computational Intelligence and Data Mining (CIDM),* Nashville, USA, 324-331, doi: 10.1109/CIDM.2009.4938667.

[26] Liu, X.-Y., Wu, J., Zhou, Z.-H. (2009). Exploratory undersampling for class-imbalance learning, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 39, No. 2, 539-550, doi: 10.1109/TSMCB.2008.2007853.

[27] Lundberg, S.M., Lee, S.-I. (2017). A unified approach to interpreting model predictions, In: *Proceedings of the 31st International Conference on Neural Information Processing Systems,* Long Beach, USA, 4766-4777.

[28] McCann, M., Johnston, A. (2008). SECOM dataset UCI machine learning repository, from *https://archive.ics.uci.edu/dataset/179/secom*, accessed October 20, 2024.

[29] Pruvost, G., Derbel, B., Liefooghe, A., Li, K., Zhang, Q. (2020). On the combined impact of population size and sub-problem selection in MOEA/D, In: Paquete, L., Zarges, C. (eds.), *Evolutionary computation in combinatorial optimization*, *EvoCOP 2020, Lecture notes in computer science*, Vol 12102, Springer, Cham, Switzerland, 131-147, doi: 10.1007/978-3-030-43680-3_9.

[30] Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD),* Anchorage, USA, 2623-2631, doi: 10.1145/3292500.3330701.

[31] Moldovan, D., Anghel, I., Cioara, T., Salomie, I. (2020). Particle swarm optimization based deep learning ensemble for manufacturing processes. In: *Proceedings of 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP),* Cluj-Napoca, Romania, 563-570, doi: 10.1109/iccp51029.2020.9266269.
[32] Chazhoor, A., Mounika, Y., Vergin Raja Sarobin, M., Sanjana, M.V., Yasashvini, R. (2020). Predictive maintenance using machine learning based classification models, *IOP Conference Series: Materials Science and Engineering*, Vol. 954, No. 1, Article No. 012001, doi: 10.1088/1757-899X/954/1/012001.
[33] Nguyen, D.-K., Chan, C.-L., Phan, D.-V. (2022). Gradient deep learning boosting and its application on the imbalanced datasets containing noises in manufacturing, In: *Proceedings of 2021 International Conference on Security and Information Technologies with AI, Internet Computing and Big-data Applications,* Cham, Switzerland, 225-235, doi: 10.1007/978-3-031-05491-4_23.
[34] Liu, F., Yang, J., Pedrycz, W., Wu, W. (2022). A new fuzzy spiking neural network based on neuronal contribution degree, *IEEE Transactions on Fuzzy Systems*, Vol. 30, No. 7, 2665-2677, doi: 10.1109/tfuzz.2021.3090912.

APEM journal
Advances in Production Engineering & Management
Letnik 20 | Številka 4 | December 2025 | Strani 458–474
https://doi.org/10.14743/apem2025.4.552
ISSN 1854-6250
Spletna stran: apem-journal.org
Izvirni znanstveni članek

Večkriterijska izbira značilk in ansambelski pristop z učenjem v lastnem tempu za zaznavanje napak v polprevodniški proizvodnji

Zheng, H.[a], **Gao, X.**[a,*], **Yang, X.**[a], **Jing, G.**[a], **Yang, M.**[a], **Liu, Y.**[a]

[a]School of Mechanical Engineering, Xi'an University of Technology, Xi'an, P.R. China

POVZETEK

V polprevodniški proizvodnji se zaznavanje napak običajno izvaja na podlagi visokodimenzionalnih procesnih podatkov. Ti podatki pogosto izkazujejo neuravnoteženost in prekrivanje razredov, kar predstavlja pomemben izziv za doseganje zanesljive klasifikacijske učinkovitosti. Za reševanje teh težav prispevek predlaga veškriterijski pristop za izbiro značilk in ansambelsko učenje v lastnem tempu (MOFS-SPE). Predlagani pristop za izbiro značilk uporablja veškriterijski evolucijski algoritem na osnovi dekompozicije (MOEA/D). V tem postopku sta kot ciljni funkciji uporabljeni površina pod krivuljo natančnost–odziv (AUPRC) in vrednost R, s katerima se identificirajo podmnožice značilk, ki so močno povezane s kakovostnimi kazalniki. Poleg tega pristop združuje ansambelsko učenje v lastnem tempu (SPE) z drevesnimi klasifikatorji za učinkovito obravnavo neuravnoteženih in prekrivajočih se podatkov. Eksperimenti, izvedeni na realnem naboru podatkov iz polprevodniške proizvodnje (podatkovni nabor SECOM), potrjujejo učinkovitost predlaganega pristopa. V primerjavi z uporabo celotnega nabora značilk izbrane značilke povečajo površino pod krivuljo ROC (AUROC) z 0,685 na 0,770 ter AUPRC z 0,932 na 0,972. Pri uporabi SPE se specifičnost modela odločitvenega drevesa izboljša z 0,048 na 0,667, kar bistveno poveča zanesljivost prepoznavanja okvarjenih izdelkov. Predlagani pristop predstavlja uporabno referenco za inteligentno kontrolo kakovosti v proizvodnji polprevodnikov.

PODATKI O ČLANKU

Ključne besede:
Proizvodnja polprevodnikov;
Zaznavanje napak;
Kontrola kakovosti;
Neuravnoteženost razredov;
Prekrivanje razredov;
Večkriterijska izbira značilk;
Ansambelsko učenje v lastnem tempu;
Strojno učenje

Kontaktna oseba:
xaut66gxq@163.com
(Gao, X.)

Zgodovina članka:
Prejet 27. avgusta 2025
Popravljen 30. novembra 2025
Sprejet 3. decembra 2025