Prediction of Cross-Sectional Images and Proposing Processes with Neural Network

Kohei Motojima, Hayato Sugiyama, Kaede Ameyama, Chiho Ueta
Taiyo Holdings Co., Ltd.
388 Ohkura, Ranzan-machi, Hiki-gun, Saitama 355-0222, Japan
Ph: 81-493-62-7777; Fax: 81-493-62-2330
Email:motojima.kohei.vc@taiyo-hd.co.jp

Abstract
We developed a novel system that quickly predicts cross-sectional images from experimental conditions to minimize cross-sectional processing. This has become a bottleneck in observing pattern shapes of materials used in substrate manufacturing. This system can predict cross-sectional images from experimental conditions and propose process conditions to form the desired pattern shape. We consider that using this system will improve the development speed of overall semiconductor packages.

Key words
neural network, machine learning, simulation, cross-sectional images, efficiency

I. Introduction
Recently, with the development of artificial intelligence (AI) and Big Data, semiconductor packages are required to have high computation capability. To meet this demand, semiconductors are miniaturized, and the semiconductor packages are rapidly becoming denser and larger due to the development of chiplet technology and heterogeneous technology. The connection method between the chip and the substrate changes based on the number of I/Os on the chip. Accordingly, the materials used for package substrates must offer functionality and processing accuracy to ensure reliable bonding with various materials. The pattern shapes of the functional materials need to be controlled because it is known that pattern shapes significantly impact the long-term reliability of semiconductor packages. To control the pattern shape, it is important to design materials and process technology, and this accuracy requirement is becoming stricter every year [1]. Therefore, whenever a new semiconductor package is developed, many substrate prototypes are fabricated and then subjected to destructive cross-section observations. This process is crucial for the material design and establishing process technology [2]. Additionally, small-diameter vias can be opened due to the evolution of lithography equipment and laser equipment. As a result, there are increasing demands for pattern shape control, such as residuals and taper angles at the bottom of the vias. Interlayer insulations and solder resist (SR) applied high-density designs for a wide variety of semiconductor packages with especially demanding pattern shapes. Because they require different film thicknesses, pattern shapes, and via sizes. To meet these demands, the cross-sections must be polished so that the center of the via diameter can be observed about the combination of film thickness and via size for each condition. Observations of the center cross-section of a via of several to several ten micrometers require advanced polishing techniques and significant experimental costs. Currently, no equipment can efficiently observe, so it can take several days to several weeks to complete a task that doesn't require a human resource. Therefore, it is a problem that not enough time can be allocated to creative work in these days of manpower shortages. For the future evolution of semiconductor packaging, new technologies are needed to efficiently and quickly achieve the development of revolutionary functional materials and the establishment of mass production process technologies, along with advances in equipment.
In the field of semiconductor resist, pattern shapes are predicted in advance using simulation tools based on theoretical calculations, such as PROLITH [3]. There are essential tools in the semiconductor manufacturing industry. However, simulations like PROLITH don't exist for back-end packaging materials such as SR. Because it is difficult to construct theories for functional materials with increasingly complex compositions to achieve the required physical properties. One of the factors that make theory construction more difficult is the dependence on individual expertise in process technology.

Regression analysis is an effective technique for dealing with phenomena that are difficult to theorize. It uses past experimental data to construct a statistical model that predicts experimental results from experimental conditions. The statistical model can predict each result from past experimental results even for complex functional materials for which the theory has not been understood, such as packaging materials. However, the output of common regression analysis methods such as the least absolute shrinkage and selection operator (LASSO) [4] and the Gaussian process (GP) [5] are one-dimensional. In the case of multi-dimensional experimental results such as images, the computational cost is very high because a separate model must be constructed for each dimension. Artificial neural network (ANN) [6] is a method to predict the results of multi-dimensional with a single model. However, ANN has many parameters and requires a large amount of data to construct the model. It is difficult to apply packaging materials that require small-quantity and large-variety products. Therefore, there is a need for new technology that can be applied to packaging materials.

The objective of this study is to provide a novel system for efficient materials design and process technology establishment by minimizing the cross-sectional observations of packaging materials. This system can predict images captured by an electron microscope from experimental conditions and propose process conditions to form the desired pattern shape of the functional material. This system has two functions. One function is to predict the cross-sectional images of pattern shapes from experimental conditions, and the other is to propose process conditions which are functional materials that form the desired pattern shapes. In this study, the functions of each were verified, and it was confirmed that the experimental cost was reduced, and the prediction accuracy was ensured.

II. Method

A. Algorithm

In this study, we propose a novel image prediction system that combines regression analysis methods such as LASSO and GP with a variational autoencoder (VAE) [7], which is used in image generation AI. VAE is a type of deep neural network consisting of an encoder and a decoder. Through the encoder, the input data are compressed into numbers of arbitrary dimensions called latent variables. The latent variables are then input to the decoder to reconstruct the input data. Latent variables are generated based on normal distributions consisting of the average and variance calculated by the encoder, so VAE is a good method to generate data. The algorithm of this system is shown in Fig. 1. This system consists of a VAE part and a regression analysis (RA) part. The VAE part deals with cross-sectional images. The cross-sectional images don't need to be associated with experimental conditions. Therefore, it is easy to prepare a large amount of data, including data collected in the past. Experimental conditions consist of material information representing the characteristics of packaging materials and process conditions such as exposure, development, and heat curing temperatures. The VAE model is constructed to compress cross-sectional images into latent variables of arbitrary dimensions. Here, we use "encoder" to include the process of generating latent variables. The RA part deals with cross-sectional images associated with experimental conditions from past experimental data. First, the cross-sectional images are input to the encoder to obtain the latent variables $\boldsymbol{Z}$. Next, RA models ($\boldsymbol{Z} = f(\boldsymbol{X})$) are constructed to predict the latent variables from the experimental conditions $\boldsymbol{X}$. By inputting $\boldsymbol{X}$ to the RA models, $\boldsymbol{Z}$ could be predicted, and by inputting predicted $\boldsymbol{Z}$ to the decoder, cross-sectional images could be predicted. Therefore, this system makes it possible to predict pattern shapes, which was not possible in the past, for packaging materials, which require small-quantity and large-variety products for which it is difficult to construct a theory.

B. Process Optimization

This system can not only predict the cross-sectional image of pattern shapes from experimental conditions but also propose process conditions based on the desired pattern shapes. First, a desired cross-sectional image is prepared. Next, many virtual experimental conditions are generated by assigning process conditions. The virtual experimental conditions are input into the RA models, latent variables are predicted and then input into the decoder to predict the cross-sectional images of the virtual experimental conditions. The similarities between the desired image and the generated images are calculated, and the process condition of any experimental condition with a high similarity can be proposed as the process condition that forms the desired cross-sectional image. The similarity is calculated using the structural similarity (SSIM) [8], which is often used for image comparison.

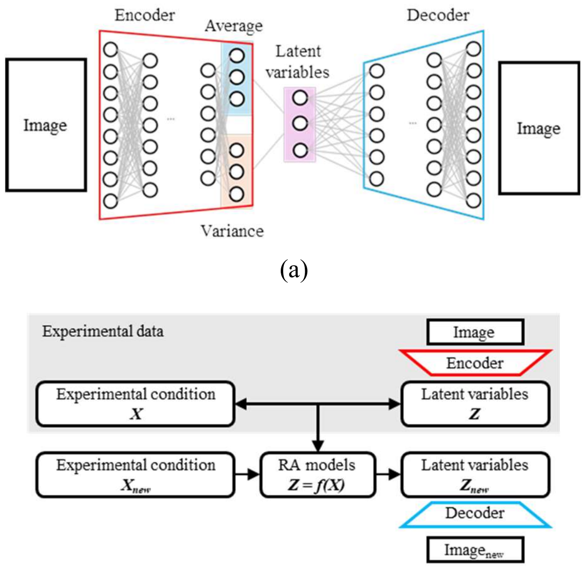


Fig. 1. Algorithm overview: (a) VAE part and (b) RA part.

III. Result and Discussion

A. Case Study

To demonstrate the effectiveness of this system, cross-sectional images of dry film type SR were dealt with. SR is a negative photosensitive and functional material related to the reliability of the connection between the chip and the substrate. It has many physical properties that are particularly required, including correlation with other materials used in packaging materials. Therefore, they are complex composite materials consisting of numerous organic and inorganic components. The main processes for forming pattern shapes are exposure and development. In the case of exposure, factors affecting the pattern shape are light scattering and absorption unique to composite materials. The main development method is the spraying of alkaline solution, and in addition to chemical factors such as the adhesion strength of SR depending on the base substrate and the swelling process depending on the type of alkaline solution, physical factors such as spray pressure are also considered. Since a complex combination of chemistry and physics must be considered to construct a theory, no effective theory of pattern shape derivation has been established. From the viewpoint of mass production using a variety of equipment, SR requires a wide process margin. Factors such as material type and film thickness, combined with complex process conditions, increase the number of targets for cross-sectional observation. Therefore, it takes several days to several weeks to observe a cross-section of the pattern shape of one objective examination. This example verification deals with a cross-sectional image of phi 70μm via taken with a scanning electron microscope (SEM).

B. VAE Part

In the VAE part, the VAE model was constructed. The SEM images were of various sizes, but all were converted to 300 pixels x 300 pixels to standardize the dimension of the input layer. The encoder had 15 layers, and the decoder had 13 layers. The latent variables were set to 5 dimensions. If the number of dimensions of latent variables is too large, the number of RA models will increase too much. If the number of dimensions of latent variables is too small, the reconstruction performance will worsen. Therefore, 5 dimensions were determined to be the most appropriate. The number of training images is 4368. The input images and reconstructed images by inputting the latent variables generated by the encoder to the decoder are shown in Fig. 2. None of the images were clearly reconstructed. Therefore, even if latent variables can be accurately predicted, there is a risk that the images will be blurred, and pattern shapes will not be recognized. The training images contain polishing marks and noise that should not be reconstructed. There are images of various contrasts because different researchers have different contrasts. In addition, some substrate types, such as those containing glass cloth, are not shown in other images. We consider that the accuracy of the pattern shapes of SR was poor because of the attempt to reconstruct the pattern shapes including the unimportant part.

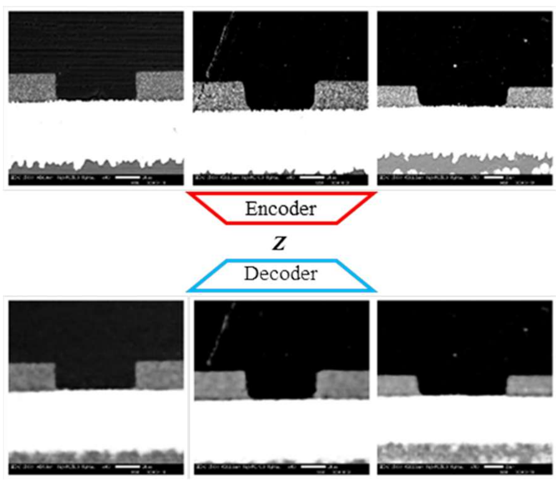


Fig. 2. Results of VAE model. Top images indicate input images and under images indicate reconstructed images.

To improve the accuracy of the reconstruction, only the SEM image information below the image, the base material, and SR were made white, and the other areas were made black. The addition of processing is expected to eliminate noise and the dependence on the researcher. Next, the VAE model was reconstructed using processed images, under the same model construction conditions and the same amount of data. The input images and reconstructed images by inputting latent variables generated by the encoder into the decoder are shown in Fig. 3. Compared to Fig. 2, the accuracy of the reconstruction is improved, which is attributed to the fact that the SEM images were processed in advance to eliminate noise and standardize the contrast of the image data. The 5-dimensional latent variables that represent the characteristics of the data generated are Z0, Z1, Z2, Z3, and Z4, respectively.

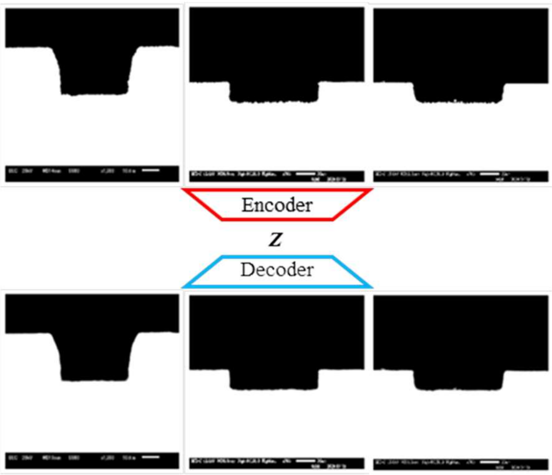


Fig. 3. Results of VAE model after image processing. Top images indicate input images and under images indicate reconstructed images.

C. RA Part

Since we were able to construct the VAE model with good reconstruction accuracy, we constructed RA models to predict each latent variable. 129 samples of cross-sectional images associated with experimental conditions were used to construct the RA models. Experimental conditions consist of 404 variables: absorbance spectrum of SR, absorbance spectrum of PET, reflectance spectrum of substrate, intensity spectrum of exposure (assuming LED light source), exposure dose, defocus, development time, and water rinse time. Process condition variables include intensity spectrum of exposure, exposure dose, defocus, development time, and rinse time. Each spectrum is shown in Fig. 4. The absorbance, reflectance, and intensity at each wavelength are handled as input variables. The 129 samples were divided into 90 samples of training data and 39 samples of test data. Each RA model using training data was constructed using multiple regression analysis methods, and the regression analysis method with the highest prediction accuracy for the test data was considered the optimal regression analysis method. The regression analysis methods tested were partial least squares (PLS) [9], ridge regression (RR) [10], LASSO, elastic net (EN) [11], support vector regression (SVR) [12], random forest (RF) [13], extreme gradient boosting (XGB) [14], and GP. As preprocessing of the data, spectra were derived and smoothed by the Savitzky-Golay method [15], and the whole data were autoscaling. Table. I shows the optimal regression analysis method, the coefficient of determination (r^2), and the root mean squared error (RMSE) for each latent variable. The actual values vs. the predicted values for each latent variable are shown in Fig. 5. From Fig. 5, the latent variables other than Z3 are close to the diagonal of the test data and can be predicted from the process conditions. This indicates that the relationship between the experimental conditions and Z3 cannot be established. Therefore, we investigated the effect of Z3 on the image by generating Z3 with ±0.24 deviation from the Z3 of the original data from $RMSE_{test}$. The result is shown in Fig. 6. There was no clear change in the image generated. We consider that Z3 is used for detailed reconstruction, such as noise removal. Therefore, this is not considered a significant issue even if the relationship between the experimental conditions and the Z3 value is not expressed in this case. Fig. 7 shows the predicted images by inputting the predicted latent variables of the test data into the decoder. Because the VAE model and RA models were both constructed with good accuracy, the pattern shapes were also predicted with good accuracy. It takes about 40 seconds from the input of the experimental conditions to the output of the images, which means that each cross-sectional image can be predicted in about 1 second. In this way, it was confirmed that the accuracy of the prediction and the experimental cost were effective.

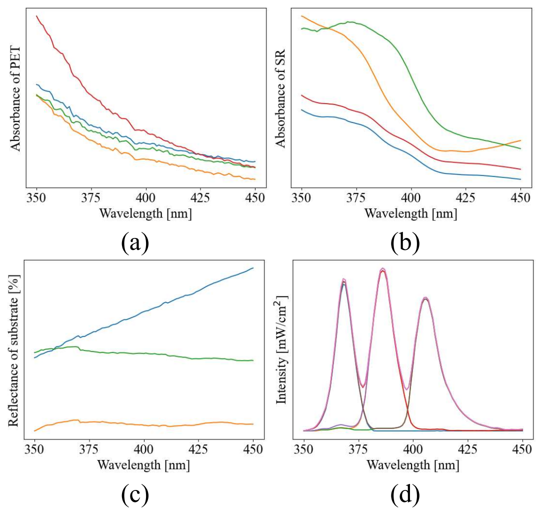


Fig. 4. Spectra: (a) absorbance of PET, (b) absorbance of SR, (c) reflectance of substrate, (d) intensity of exposure.

Table. I. Prediction accuracy

Z	Regression analysis	r^2_{train}	r^2_{test}	$RMSE_{train}$	$RMSE_{test}$
Z0	RF	0.899	0.407	0.311	0.653
Z1	GP	0.994	0.437	0.057	0.513
Z2	GP	0.939	0.902	0.368	0.540
Z3	EN	0.539	0.456	0.478	0.422
Z4	GP	0.847	0.267	0.358	0.400

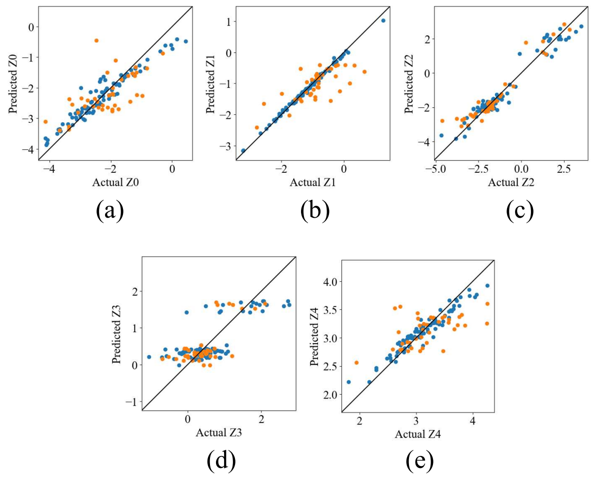


Fig. 5. The actual values vs. the predicted values for each latent variable: (a) Z0, (b) Z1, (c) Z2, (d) Z3, (e) Z4. Blue points indicate training data and orange points indicate test data.

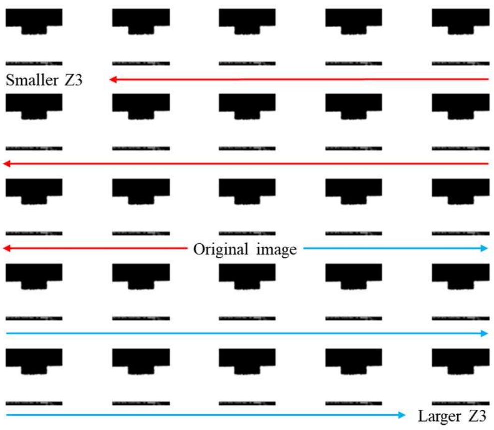


Fig. 6. Effect of Z3 on the image. The original image is in the center, and the Z3 value is smaller toward the red arrows and larger toward the blue arrows.

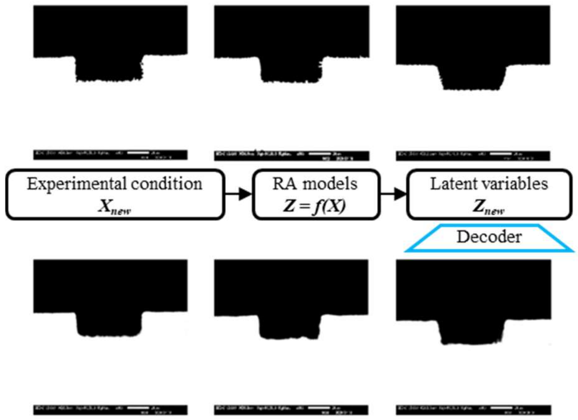
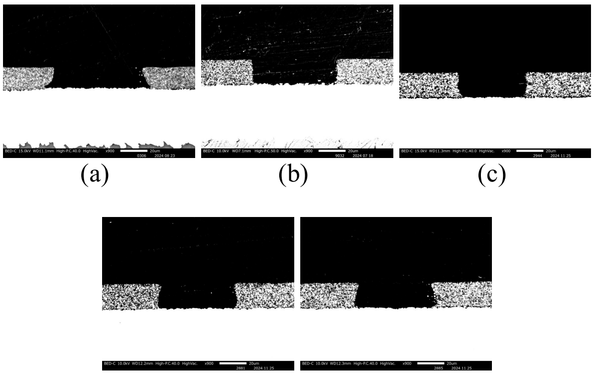


Fig. 7. Predicted images. Top images indicate actual images and under images indicate predicted images.

D. *Process Optimization*

We attempted to propose process conditions to form the desired pattern shape. The subject was a specific functional material, and an undercut was observed in all experimental results as shown in Fig. 8 (a) when light with a distribution centered at the exposure wavelength of 365 nm was applied under the exposure process conditions. Therefore, the target was to propose process conditions that would prevent undercuts. The target pattern shape is shown in Fig. 8 (b). First, the target pattern shape was processed by image processing. Next, 129 cross-sectional images associated with the experimental conditions were used as training data to construct RA models based on Table I. Next, about 1000 samples of virtual experimental conditions were generated by varying the exposure wavelength, exposure dose, and development time. The latent variables predicted by RA models were input into the decoder and SSIM was calculated after predicting the cross-sectional images. Experiments were conducted and verified based on three high SSIM conditions. The cross-sectional images of the proposed experimental conditions are shown in Fig. 8 (c-e), and the pattern shape values are shown in Table. II. The numbers in brackets are the errors from (b). Compared to Fig. 8 (a), undercutting was controlled overall. Fig. 8 (b) and Fig. 8 (c) are similar in pattern shape, but there is an error in numerical values. However, we consider that this opening ratio can be adjusted by the settings of the exposure machine, such as the offset value. Therefore, we believe that when we search for conditions around the proposed experimental conditions, the numerical values will be approached. Since it takes only about 30 seconds to make proposed process conditions from 1,000 samples, its effectiveness can be confirmed by condition screening, and it is expected to contribute to the reduction of experimental costs.



(a) (b) (c)

(d) (e)

Fig. 8. Pattern shape: (a) subject and (b) target, (c) optimal condition 1, (d) optimal condition 2, (e) optimal condition 3.

Table. II. Optimization of processes

	Top [µm]	Middle [µm]	Bottom [µm]
Subject (a)	66	67	76
Target (b)	64	63	60
(c)	51(-22%)	51(-19%)	47(-21%)
(d)	62(-3%)	58(7%)	58(-3%)
(e)	53(-17%)	58(-8%)	61(2%)

IV. Conclusion

In this study, a novel image prediction system was developed to reduce destructive cross-sectional observation, which is a large experimental cost in material design and establishment of processes for functional materials applied to semiconductor packages, and to derive optimal solutions from many experimental conditions. As a feature not found in conventional methods, the combination of regression analyses and VAE enabled highly accurate prediction of cross-sectional images and optimization of process conditions on small data. Since this system takes only a few seconds to output, screening conditions are reduced, and the development time for new semiconductor packages, which involves complex technology for materials, machines, and process matching, is expected to be shortened. In addition, it can predict phenomena that involve image judgment, such as residues, that are difficult to quantify, so it can be applied in a wide range of applications. In this study, we focused on images. In addition to this technology, we completed the construction of AI models to predict measured values of pattern shapes. By combining these, it becomes possible to discuss not only images but also numerical data. We expect that this system and the use of data across industries will contribute to the development of the semiconductor packaging market.

References

[1] K. Fujimoto, Y. Okaya, T. Tai, S. Kuramochi, "Development of glass core substrate with the stress analysis, transmission characteristics and reliability," 2024 IEEE 74th Electronic Components and Technology Conference, pp.546-550, 2024.

[2] K. Murai, H. Onozeki, D. Kang, K. Hirano, K. Mitsukura, "Study of fabrication and reliability for the extremely large 2.5D advanced package," 2023 IEEE 73rd Electronic Components and Technology Conference, pp.899-906, 2023.

[3] C. A. Mack, "PROLITH: a comprehensive optical lithography model," Proc. SPIE 0538, Optical Microlithography, pp. 207-220, 1985.

[4] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society. Series B, vol. 58, no. 1, pp. 267-288, 1996.

[5] C. E. Rasmussen, C. K. I. Williams, "Gaussian processes for machine learning", The MIT Press, 2006.

[6] A. Prieto, B. Prieto, E. M. Ortigosa, E. Ros, F. Pelayo, J. Ortega, I. Rojas, "Neural networks: An overview of early research, current frameworks and new challenges," Neurocomputing, vol. 214, pp. 242-268, 2016.

[7] D. P. Kingma, M. Welling, "Auto-encoding variational bayes," arXiv:1312.6114, 2013.

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on image processing, vol. 13, no. 4, 2004.

[9] S. Wold, S. Sjöstöm and L. Eriksson, "PLS-regression: a basic tool of chemometrics," Chemometrics and Intelligent Laboratory Systems, vol. 58, no. 2, pp. 109-130, 2001.

[10] A. E. Hoerl, R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," Technometrics, vol. 12, no. 1, pp. 55-67, 1970.

[11] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society B, vol. 67, no. 2, pp. 301-320, 2005.

[12] A. J. Smora and B. Schoelkopf, "A tutorial on support vector regression," Statistics and Computing, vol. 14, no. 3, pp. 199-222, 2004.

[13] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-35, 2001.

[14] W. Zhang, C. Wu, H. Zhong, Y. Li, L. Wang, "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization," Geoscience Frontiers, vol. 12, no. 1, pp. 469-477, 2021.

[15] A. Savitzky, M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," Analytical Chemistry vol. 36, no. 8, pp. 1627-1639, 1964.