**ASTEEC Conference Proceeding: Computer Science**
3rd International Conference on Information Science and Technology Innovation (ICoSTEC)
July 27, 2024, Yogyakarta, Indonesia

# Application of Support Vector Machine (SVM) for Occupancy Status Data of Rehabilitation and Reconstruction Houses Post-Mount Merapi Eruption

Nurhadi Wijaya[1], Sugeng Winardi[2], Kasmawaru[3], Anang F. Rahman[4], Yazid Irfan[5]

*Informatics Department, Information System Department, Universitas Respati Yogyakarta, Universitas Dipa Makassar, Universitas Madura*

[1]nurhadi@respati.ac.id
[2]sugeng@respati.ac.id
[3]kasmawaru@undipa.ac.id
[4]anang@unira.ac.id
[3]20220039@Respati.ac.id

***Abstract*— This research focuses on the application of the Support Vector Machine (SVM) classification algorithm to analyze the occupancy status data of rehabilitation and reconstruction houses (Rehab-Rekon) post-Merapi eruption. The eruption caused significant damage to residential areas, necessitating urgent rehabilitation and reconstruction efforts. The study aims to develop a predictive model that can classify the occupancy status of houses post-rehabilitation and reconstruction, aiding decision-making in aid distribution and reconstruction prioritization. The SVM algorithm was chosen for its ability to handle complex data and generalize well, thus improving the accuracy of occupancy status predictions. The dataset includes 2,516 houses, with 2,146 occupied and 370 with unclear occupancy status. The research seeks to validate the hypothesis that the SVM algorithm can effectively classify the occupancy status of Rehab-Rekon houses with satisfactory accuracy 91,27%. The findings are expected to demonstrate the potential of SVM in disaster recovery, particularly in the rehabilitation and reconstruction sector, providing valuable insights for stakeholders in planning and executing targeted and sustainable rehab-rekon programs.**

***Keywords*— Support Vector Machine (SVM), Rehabilitation and Reconstruction, Post-Eruption of Mount Merapi, Data Classification, Home Occupancy Status**

## I. Introduction

The eruption of Mount Merapi that occurred in 2010 caused major damage to surrounding residential areas, triggering an urgent need for rehabilitation and reconstruction of houses for eruption victims. To overcome the impact of this disaster, various rehabilitation and reconstruction programs have been implemented by the government and related institutions[1][3]. One indicator of the success of this program is the residential status of houses that have been rehabilitated and reconstructed[1]. The more houses are occupied, the better the performance of the rehabilitation and reconstruction program.[2]

This research focuses on the application of the Support Vector Machine (SVM) algorithm to analyze data on the occupancy status of rehabilitation and reconstruction houses after the eruption of Mount Merapi. The SVM algorithm was chosen because of its ability to handle complex data and good generalization capabilities [4], so it can increase the accuracy of predicting house occupancy status. The data used in this research was obtained through the Rekon Rehab Satker, which has recorded data on permanent residential status. A total of 2,146 houses have been occupied and as many as 370 houses have no clear status as to whether they are occupied or unoccupied, with a total of 2,516 houses.

This research aims to develop a predictive model that can classify the occupancy status of houses after rehabilitation and reconstruction, to support decision making in distributing aid and prioritizing reconstruction in the future. By using the SVM algorithm, it is hoped that the model developed can help the government and related institutions in identifying houses that require high priority in the rehabilitation and reconstruction process, so that resource allocation can be carried out more effectively and efficiently.

In addition, this research also discusses several challenges in implementing the SVM algorithm, including selecting appropriate parameters and handling imbalanced data. Solutions to these challenges, such as the use of cross-validation techniques and sampling methods, will also be discussed to improve model performance. This research makes an important contribution to the field of disaster management, especially in the aspect of post-disaster recovery, by integrating machine learning techniques to strengthen data-based decision-making processes.

Support Vector Machine (SVM) is a powerful supervised learning algorithm used for classification and regression tasks. SVM works by finding the hyperplane that best separates data into different classes. This hyperplane maximizes the margin between the classes, ensuring robust classification even with high-dimensional data. One key feature of SVM is the use of kernel functions, such as the radial basis function (RBF), which allows it to handle non-linear data by mapping input features into higher-dimensional spaces where a linear separation is possible. SVM has been widely applied in various fields, including bioinformatics for disease classification, image recognition, and text categorization. Its ability to manage large feature spaces and provide high accuracy makes it a preferred choice for many practical applications. However, the selection of an appropriate kernel and tuning of parameters such as the regularization parameter and kernel parameters are crucial for optimal performance[4][7]

Research on the application of the Support Vector Machine (SVM) classification algorithm for data on the occupancy status of rehabilitation and reconstruction houses after the eruption of Mount Merapi offers an innovative approach that supports efficiency and effectiveness in rehabilitation and reconstruction efforts. With accurate predictive capabilities, this model is expected to be a valuable tool for stakeholders in planning and implementing rehabilitation and reconstruction programs in a more focused and sustainable manner.

In a broader context, this research also shows the potential for applying the SVM algorithm in the natural disaster sector, especially in aspects of post-disaster rehabilitation and reconstruction. The model developed can help the government and related institutions in identifying houses that require high priority in the rehabilitation and reconstruction process, so that resource allocation can be carried out more effectively and efficiently. In addition, this research also discusses several challenges in implementing the SVM algorithm, including selecting appropriate parameters and handling imbalanced data. Solutions to these challenges, such as the use of cross-validation techniques and sampling methods, will also be discussed to improve model performance.

This research makes an important contribution to the field of disaster management, especially in the aspect of post-disaster recovery, by integrating machine learning techniques to strengthen data-based decision-making processes. Research on the application of the Support Vector Machine (SVM) classification algorithm for data on the occupancy status of rehabilitation and reconstruction houses after the eruption of Mount Merapi offers an innovative approach that supports efficiency and effectiveness in rehabilitation and reconstruction efforts. With accurate predictive capabilities, this model is expected to be a valuable tool for stakeholders in planning and implementing rehabilitation and reconstruction programs in a more focused and sustainable manner.

## II. RELATED WORKS

Algorithm Support Vector Machine (SVM), has gained significant attention in various fields, including disaster management and rehabilitation. SVM is a supervised learning model that analyzes data for classification and regression analysis. Its ability to handle high-dimensional data and provide robust classification results makes it a suitable choice for complex datasets, such as those encountered in post-disaster scenarios.

Several studies have explored the use of SVM in disaster management. For instance, Yu et al. [5] demonstrated the potential of SVM in classifying individuals with and without common diseases, such as diabetes and pre-diabetes, using data from the National Health and Nutrition Examination Survey (NHANES). The study highlighted the effectiveness of SVM in handling high-dimensional data and achieving high classification accuracy. Similarly, Guido et al. [6] provided an extensive review of SVM applications in healthcare, emphasizing its role in diagnosis, prognosis, and disease outcome prediction. The review also discussed various SVM-based models developed to enhance generalization capabilities and improve performance metrics.

In the context of disaster management, SVM has been employed for various predictive modeling tasks. A study by Naceur[7] focused on developing a landslide susceptibility assessment model using SVM, refined through the Bayesian algorithm for hyperparameter optimization. The model demonstrated robust performance in predicting landslide susceptibility in the Qinghai-Tibet Plateau, highlighting the applicability of SVM in geospatial and environmental data analysis. Another study by Ahmad[8] introduced an improved short-term power load forecast model using SVM, optimized with Particle Swarm Optimization (PSO) parameters. This model showcased the versatility of SVM in handling different types of data and improving prediction accuracy.

The application of SVM in post-disaster rehabilitation and reconstruction is particularly relevant given the need for accurate and efficient classification of housing status. Previous research has shown the potential of various machine learning algorithms in this domain. For example, Wijaya et al. [9] applied the Naive Bayes algorithm to classify the occupancy status of houses rehabilitated and reconstructed after the Merapi eruption. The study achieved an accuracy of 89.59%, demonstrating the feasibility of using machine learning for post-disaster housing classification. Subsequent studies by Wijaya and colleagues explored the combination of Naive Bayes with Chi-Square feature selection and the use of Decision Tree C.45[10] both yielding promising results in terms of classification accuracy.

Recent advancements have also seen the application of more sophisticated algorithms. In 2023, Wijaya and Mustiadi [11] employed the Adaboost algorithm for classifying post-disaster housing status, achieving high accuracy and demonstrating the continuous improvement in machine learning techniques for disaster management. These studies underscore the importance of selecting appropriate algorithms and optimizing their parameters to enhance classification performance.

The current study aims to build on this body of work by applying SVM to classify the occupancy status of houses rehabilitated and reconstructed after the Merapi eruption. SVM's ability to handle complex and high-dimensional data makes it a suitable candidate for this task. The study will leverage data collected by the Satker Rehab Rekon, which includes detailed records of housing status post-rehabilitation and reconstruction. By applying SVM, the study aims to develop a predictive model that can accurately classify housing status, thereby supporting decision-making in the allocation of resources and prioritization of reconstruction efforts.

In conclusion, the application of SVM in post-disaster rehabilitation and reconstruction represents a promising approach to improving the efficiency and effectiveness of disaster management efforts. The existing literature highlights the versatility and robustness of SVM in handling various types of data and achieving high classification accuracy. This study will contribute to the growing body of knowledge by demonstrating the applicability of SVM in classifying post-disaster housing status, ultimately aiding in the efficient allocation of resources and prioritization of reconstruction efforts.

## III. METHOD

This research uses a quantitative approach with experimental methods to test the application of the Support Vector Machine (SVM) Classification Algorithm in classifying data on the occupancy status of rehabilitation and reconstruction houses after the eruption of Mount Merapi. This research was carried out through several stages, namely data collection, data preprocessing, implementation of the SVM algorithm, model evaluation, and result validation.

The data used in this research is data on the occupancy status of rehabilitation and reconstruction houses after the eruption of Mount Merapi obtained from the Rekon-Rehab Working Unit. This data includes information about 2,516 houses, of which 2,146 houses are occupied and 370 houses have unclear occupancy status. This data was collected through field surveys and administrative data recorded by the Rekon Rehab Satker.

The data preprocessing stage involves several important steps to ensure the data is ready for use in analysis. These steps include data cleaning, data transformation, and data normalization. Data cleaning was performed to remove missing or invalid values. Data transformation is carried out to change data into a format suitable for analysis. Data normalization is carried out to ensure that all features have the same scale, so that the SVM algorithm can work more effectively.

Once the data is ready, the next step is to implement the SVM algorithm to classify house occupancy status. The SVM algorithm was chosen because of its ability to handle complex data and good generalization capabilities. The implementation of the SVM algorithm was carried out using the Scikit-Learn library in Python, which is a popular tool for machine learning. The processed dataset is divided into a training set (80% of the total data) and a testing set (20% of the total data). This division provides significant training for the SVM model, while ensuring an appropriate test set to validate the model's predictions.

The SVM model that has been developed is evaluated using evaluation metrics such as accuracy, precision, recall, and F1-score. Accuracy measures how often the model makes correct predictions. Precision measures how many positive predictions are actually positive. Recall measures how much positive data the model managed to predict correctly. F1-score is the harmonic average of precision and recall, which provides an overall picture of model performance. Additionally, ROC and AUC curves are also used to measure model performance. The ROC curve shows the model's ability to differentiate between positive and negative classes, while AUC measures the probability of output from a randomly selected sample from the positive or negative population.

Validation of results is carried out using test data that has been previously separated. The model prediction results are compared with actual data to measure model performance. In addition, cross-validation techniques are also used to ensure that the model is not overfitting and can work well on data that has never been seen before. The cross-validation technique involves dividing the data into subsets, where each subset is used as testing data while the other subset is used as training data. This process is repeated several times to get a more accurate picture of the model's performance.

After the model has been evaluated and validated, the next step is to analyze and interpret the results obtained. This analysis involves examining evaluation metrics to determine how well the model performs in classifying home occupancy status. In addition, analysis is also carried out to identify patterns or trends in the data that can provide further insight into the factors that influence home occupancy status. The results of this analysis are then used to make recommendations for the government and related institutions in identifying houses that require high priority in the rehabilitation and reconstruction process. The method used can be shown in the following image below

.

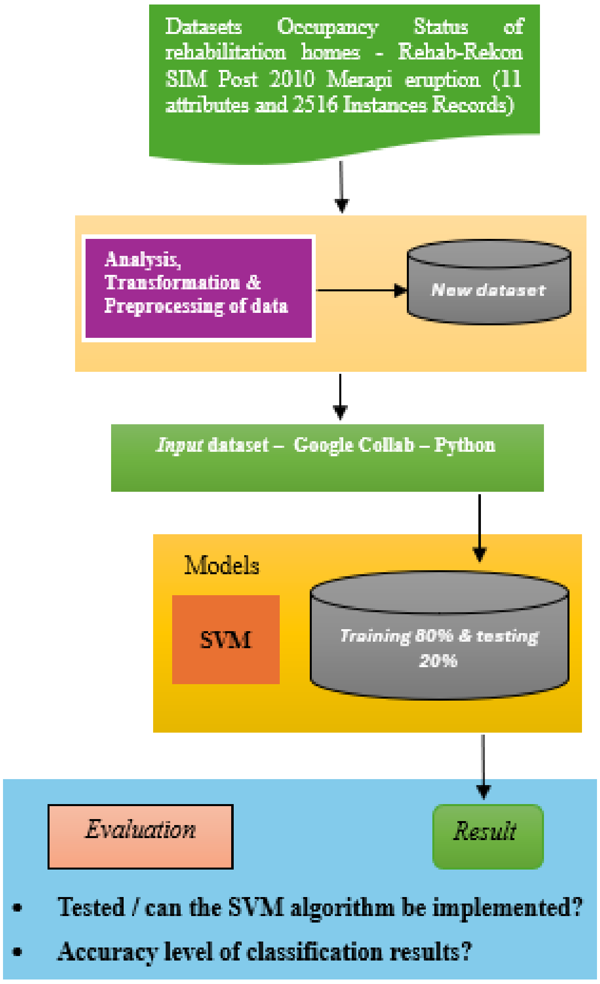


Fig. 1, Method for implementing Support Vector Machine (SVM) Classification Algorithm in classifying data on the occupancy status of rehabilitation and reconstruction houses after the eruption of Mount Merapi

## IV. RESULTS

The results of this research show that the SVM algorithm was successfully applied in classifying the occupancy status of Rehab-Rekon houses with a satisfactory accuracy of 91.27%. The model developed can help identify houses that require high priority in the rehabilitation and reconstruction process, especially regarding whether the house is occupied or not, so that resource allocation can be carried out more effectively and efficiently. In addition, this research also discusses several challenges in implementing the SVM algorithm, including selecting appropriate parameters and handling imbalanced data. Solutions to these challenges include the use of cross-validation techniques and sampling methods to improve model performance.

This research contributes to the field of disaster management, especially in the aspect of post-disaster recovery, by integrating machine learning techniques to strengthen data-based decision-making processes. Research on the application of the Support Vector Machine (SVM) Classification Algorithm for data on the occupancy status of Rehab-Rekon houses after the eruption of Mount Merapi offers an innovative approach that supports efficiency and effectiveness, with accurate predictive capabilities, so that it can become a valuable tool for stakeholders in planning and carry out rehabilitation and reconstruction programs in a more targeted and sustainable manner.

This research also shows that the SVM algorithm can minimize empirical structural risk, which is a combination of two components: empirical loss and model complexity. Empirical loss measures how well the model fits the training data, while model complexity measures the complexity of the discovered hyperplane. SVM looks for a hyperplane that has a maximum margin, namely the shortest distance between the hyperplane and the closest data points from both classes.

Model evaluation is carried out using metrics such as confusion matrix and ROC curve to measure classification performance and accuracy. The ROC curve results can be shown in the following figure :

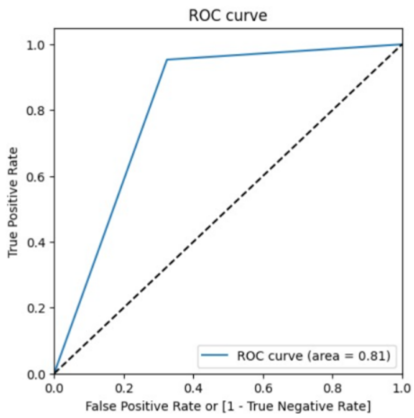


Fig. 2, Results of ROC Curve – SVM on Occupany House Status

This process is part of an effort to develop predictive models that can support decision making in post-disaster rehabilitation and reconstruction. The evaluation results show that the SVM model has high accuracy in classifying home occupancy status, with an accuracy value of more than 91.27%. Additionally, this model also shows good performance in other evaluation metrics such as 94% precision, 95% recall, and 94% F1-score as shown in figure 3 below :

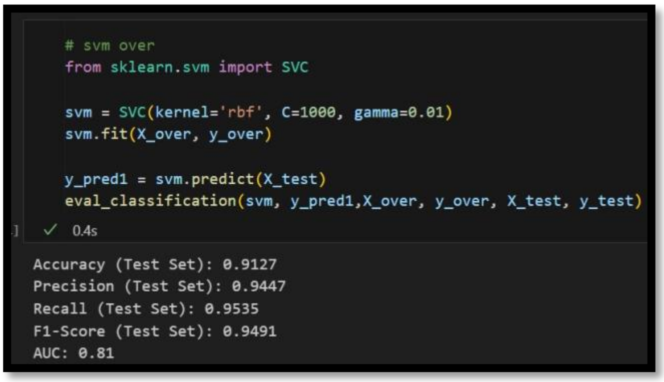


Fig. 3, Results of SVM on Occupany House Status

Overall, this research succeeded in showing that the SVM algorithm can be applied well in classifying occupancy status data for rehabilitation and reconstruction houses after the eruption of Mount Merapi. The model developed can be an effective tool in supporting data-based decision making, thereby increasing efficiency and effectiveness in post-disaster rehabilitation and reconstruction efforts. This research also makes an important contribution to the field of disaster management, by integrating machine learning techniques to strengthen data-based decision-making processes

## V. Conclusions

This research shows that the Support Vector Machine (SVM) algorithm can be applied effectively to classify the occupancy status of rehabilitation and reconstruction houses after the eruption of Mount Merapi. By using data obtained from the Rehab Rekon Working Unit, the SVM model succeeded in achieving satisfactory accuracy in predicting home occupancy status. The results of this research provide an important contribution to the field of disaster management, especially in the aspect of post-disaster rehabilitation and reconstruction, by integrating machine learning techniques to strengthen data-based decision-making processes. The model developed can help the government and related institutions in identifying houses that require high priority in the rehabilitation and reconstruction process, so that resource allocation can be carried out more effectively and efficiently.

## Acknowledgment

We would like to express our sincere gratitude to all those who have supported and contributed to this research. Special thanks to the Universitas Respati Yogyakarta for providing the necessary resources and funding through the Internal Grant Program. We also extend our appreciation to the Rehabilitation and Reconstruction Task Force for their invaluable data and insights on the post-eruption housing status of Mount Merapi.

## References.

[1] B. Setiawan, Pendampingan yang Mencerahkan, 1st ed. Jakarta: Kementrian Pekerjaan Umum Direktorat Jenderal Cipta Karya Republik Indonesia, 2013.

[2] KPI-CSSRP, Key Performance Indicators: Community-based Settlements Rehabilitation and Reconstruction Project. KPI Rekompak Jateng & DIY, JRF-PSF, Status November 2014.

[3] Satker Rehabrekon, Pembangunan Permukiman Lestari: Layak huni dan Berkelanjutan. Jakarta: Satker Rehabrekon Rumah Pasca Gempa Bumi DIY & Jateng, Ditjen Cipta Karya, 2014.

[4] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009

[5] Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). “Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes.” *BMC Medical Informatics and Decision Making*, vol. 10, no. 1, pp. 1-7.

[6] Guido, R., & Schiavo, R. (2024). “Support vector machines in healthcare: A comprehensive review.” Journal of Healthcare Engineering, vol. 2024

[7] H. A. Naceur, H. G. Abdo, B. Igmoullan, et al., "Performance assessment of the landslide susceptibility modelling using the support vector machine, radial basis function network, and weight of evidence models in the N'fis river basin, Morocco," *Geosci. Lett.*, vol. 9, no. 39, 2022.

[8] Ahmad, F.A., Liu, J., Hashim, F., Samsudin, K., 2024. Short-Term Load Forecasting Utilizing a Combination Model: A Brief Review. *International Journal of Technology*. Volume 15(1), pp. 121-129

[9] W. Nurhadi, "Penerapan Algoritma Klasifikasi Naive Bayes Untuk Data Status Huni Rumah Bantuan Dana Rehabilitasi Dan Rekonstruksi Pasca Erupsi Gunung Merapi 2010," in Prosiding Seminar Nasional, Pendekatan Multidisiplin Ilmu Dalam Manajemen Bencana, Universitas Respati Yogyakarta, Daerah Istimewa Yogyakarta, vol. 1, no. 1, 2018, pp. 1-7, ISSN 2657-2397.

[10] W. Nurhadi and O. Wayan, "Evaluation of Naïve Bayes and Chi-Square Performance for Classification of Occupancy House," International Journal of Informatics and Computation (IJICOM), vol. 1, no. 2, pp. 46-54, 2019, ISSN: 2685-8711, E-ISSN: 2714-5263.

[11] N. Wijaya, M. Diqi, and I. Mustiadi, "AdaBoost Classification for Predicting Residential Habitation Status in Mount Merapi Post-Eruption Rehabilitation," Jurnal CoSciTech (Computer Science and Information Technology), vol. 4, no. 2, pp. 429-436, Aug. 2023.

[12] W. Nurhadi and E. H. Marselina, "Penerapan Algoritma Decision Tree C.45 Untuk Klasifikasi Data Status Huni Rumah Rehabilitasi Pasca Erupsi Merapi," in Prosiding Seminar Nasional, Tetap Produktif dan Eksis Selama dan Pasca Pandemi COVID-19, Universitas Respati Yogyakarta, Daerah Istimewa Yogyakarta, vol. 2, no. 1, 2020, pp. 1-7, ISSN 2657-2397.

[13] W. Nurhadi, et al., "Implementation of KNN Algorithm for Occupancy Classification of Rehabilitation Houses," International Journal of Informatics and Computation (IJICOM), vol. 4, no. 2, pp. 46-54, 2022, ISSN: 2685-8711, E-ISSN: 2714-5263.

[14] J. Han and M. Kamber, Data Mining: Concepts, Models, and Techniques. Berlin Heidelberg: Springer, 2011.

[15] F. Gorunescu, Data Mining: Concept, Models and Techniques. Craiova, Romania: Springer, 2011.

[16] L. Chen, J. Li, S. Li, X. Li, and J. Li, "An Improved KNN Algorithm for Breast Cancer Classification," Journal of Healthcare Engineering, vol. 2017, pp. 1-7, 2017.

[17] Bustami, "Penerapan Algoritma Naive Bayes untuk Mengklasifikasi Data Nasabah Asuransi," Jurnal Informatika, vol. 8, no. 1, pp. 884-898, 2014.

[18] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119-139, 1997.

[19] E. Alpaydin, Introduction to Machine Learning, 2nd ed. MIT Press, 2010.

[20] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. Springer, 2009.

[21] J. D. Kelleher and B. Tierney, Data Science: An Introduction. CRC Press, 2018.

[22] Kusrini and L. E. T., Algoritma Data Mining. Yogyakarta: Andi Offset, 2009.

[23] S. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, 2014.

[24] K. H. Kim, H. J. Kim, and I. H. Lee, "SVM-based Incremental Learning for Real-Time Spamware Detection," Expert Systems with Applications, vol. 166, p. 114150, Mar. 2021.

[25] D. V. Nguyen, T. T. Vu, V. T. Nguyen, and M. T. Thai, "A New SVM-Based Binary Classifier for Internet Traffic Classification," in 2022 IEEE Wireless Communications and Networking Conference (WCNC), Apr. 2022, pp. 1-6.

[26] Y. Zhou, L. Yang, X. Zhang, and Y. Dai, "A Hybrid Deep Learning Model for Sentiment Analysis Based on Convolutional Neural Network and Support Vector Machine," Applied Intelligence, Jan. 2022, pp. 1-3.

[27] L. Yu, K. Chen, Y. Zhang, W. Yu, and Q. Wang, "A Fast and Accurate Adaptive SVM Classification Approach Based on Deep Learning for Software Defect Prediction," Knowledge-Based Systems, vol. 113825, Feb. 2022.P. P. Ray, "Internet of things for smart homes: Technologies, security, and challenges," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, no. 1, pp. 153-164, 2017