Combining Oversampling and Pretrained Feature Extractor For Classification Diabetic Foot Uclear Thermogram Images

I Wayan Jepriana[1*], I Gede Bintang Arya Budaya[2], Gede Angga Pradipta[3], Putu Desiana Wulaning Ayu[4]

[1]Information System, Institut Teknologi dan Bisnis STIKOM Bali, Indonesia
[2]Information Technology, Institut Teknologi dan Bisnis STIKOM Bali, Indonesia
[3,4] Magister of Information System, Institut Teknologi dan Bisnis STIKOM Bali, Indonesia
*jepriana@stikom-bali.ac.id

Abstract - Diabetic Foot Ulcers (DFUs) represent a significant health concern, often leading to severe complications if not diagnosed and treated promptly. Early and accurate classification of DFUs is crucial for effective patient management. However, In the realm of machine learning, the imbalanced data problem is a prevalent issue that arises when the classes in a dataset are not represented equally. This study proposes a novel approach to enhance the classification performance of DFU thermogram images by integrating oversampling techniques with pretrained feature extractors. This study use pretrained model method with InceptionV3 architecture to automatically obtain features in the DFU thermogram datasets. Overall, InceptionV3 as a feature extractor resulted in satisfactory performance, achieving an accuracy of 83.1% on non-diabetic data and 81.1% on diabetic data. Subsequently, the second experiment incorporated the oversampling technique SMOTE, leading to an improvement in performance, with accuracy rising to 98.1% on non-diabetic data and 96.1% on diabetic data. Finally, the SMOTE IPF method achieved accuracy of 98.7%, with a precision of 99.1% for the diabetic class and 98.7% for the non-diabetic class, a recall of 98.2% for the diabetic class and 98.1% for the non-diabetic class, and F-Measure of 98.1% for both the diabetic and non-diabetic classes.

Keywords: diabetic foot uclear, SMOTE, cost sensitive learning, machine learning, imbalanced data.

I. INTRODUCTION

Diabetic foot ulcers are among the complications that can arise in people with diabetes. These ulcers can result in functional impairment, infection, and potentially even amputation. Thermography is a non-invasive diagnostic method that uses a thermal camera to measure the surface temperature of the body. Applying thermography to foot ulcers generates thermogram images, offering valuable information about temperature variations around the wound and aiding in tracking changes in the ulcers over time. For patients with foot ulcers, irregular temperature patterns or anomalies may signal tissue damage or circulatory changes linked to the wound. Thermogram images present a non-invasive means to detect temperature changes in foot ulcers. Image processing in Fig. 1 displays images of both normal and ulcerated feet.

Several studies have explored the classification of foot ulcers through machine learning and deep learning methods. These studies focus on detecting diabetic foot ulcers (DFU), extracting representative features, and employing machine learning and deep learning algorithms for segmenting and classifying DFU objects. In study [1], asymmetry features of texture and temperature between the left and right feet were utilized. Features for texture and temperature were extracted from 11 regions of interest (ROI), with asymmetry analysis performed on features from the ipsilateral and contralateral regions of the feet. Study [2] focused on extracting temperature difference patterns. Feature vectors, based on the spectrum of 3D morphological patterns and their relative positions, enabled the system to quantitatively differentiate between non-diabetic (control) and diabetic (DM) groups. In study [3], a method for extracting the Cluster Thermal Index (CTI) was proposed. The CTI measures temperature deviations between the subject and the control group, taking into account both the temperature differences between clusters and the temperature range within the control group. Study [4] utilized a frequency-based approach. Foot images were decomposed using discrete wavelet transform (DWT) and high order spectra (HOS). Various texture and entropy features were extracted.

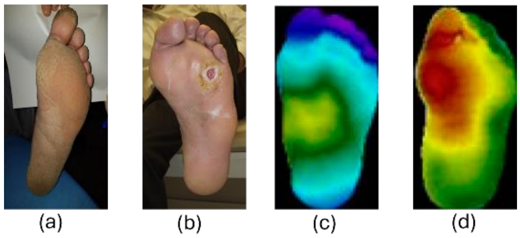

Fig. 1 (a) Image of a normal foot, (b) image of a foot ulcer, (c) thermogram of a normal foot, (d) thermogram of a foot ulcer

These combined features (DWT + HOS) were assessed using t-values and classified with the SVM method. Study [5] employed a color texture analysis approach, comparing the minimal number of color moments and dissimilarity indices. The Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) extraction methods combined with the Bag of Features (BOF) technique[6]. These methods are claimed to be quite reliable in handling image conditions that are invariant to translation, rotation, and scale transformation. Study used the average temperature to analyze both feet using the Kruskal-Wallis test. Study [7] implemented texture extraction methods, including SURF, Harris, FAST, textural, and histogram features. Other studies concentrated on Deep Learning as the primary method. For instance, study (16) employed theFuSegnet concept. A modified spatial and channel squeeze-and-excitation (scSE) module called parallel scSE or P-scSE is proposed that combines additive and max-out scSE. A new arrangement is introduced for the module by fusing it in the middle of each decoder stage. Feature extraction via convolutional neural networks (CNN) was suggested in studies [1], [2], [8]-[11], utilizing RGB image inputs and examining pixel relationships within these color spaces and the CNN showed the best feature extraction.

For developing a machine learning model, having ample training data and representative features facilitates the algorithm in identifying patterns for each class. Based on existing studies, this research focuses on addressing the problem of imbalanced data between classes in the DFU dataset. It proposes using an oversampling technique with the Safe Level SMOTE [12], [13] method to create synthetic datasets. Additionally, to improve machine learning performance, this study suggests employing cost-sensitive learning methods such as Random Forest and Naïve Bayes, which concentrate on better representing data that is more challenging to learn.

II. METHOD

The classification of diabetic foot ulcers in this study begins with splitting the image data into two parts: training data and testing data. The data split ratio is 70% for training and 30% for testing. To enhance the characteristics of the training images, image enhancement techniques such as contrast stretching, and histogram equalization are applied. Following this, each image undergoes convolution processing using a Convolutional Neural Network (CNN) method. The CNN architecture employed in this research utilizes a pretrained network model, specifically InceptionV3 [14]. InceptionV3 is a variant of the Inception model designed to improve computational efficiency and accuracy in image classification. The InceptionV3 architecture consists of several key components: the Stem, Inception blocks, Auxiliary Classifier, and final layers. At the last layer of this process, specifically at the Global Average Pooling Layer, the data dimension is reduced from (8,8,2048) to (2048) as a one-dimensional vector. The subsequent step after feature extraction is to reduce the dimensionality of these features by selecting those with high relevance to the label/class. This study uses the information gain method to measure the reduction of uncertainty or entropy in the data after knowing the value of a feature. The dimensionality reduction is tested with combinations of feature rankings, with a minimum number of five features, and evaluated based on the model's performance graph. After identifying the best features through the feature selection process, the next step is to perform data oversampling. The purpose of this process is to balance the number of data points in each label/class. In this study, there is a significant difference in the number of data points between the control group (CG) and diabetic mellitus (DM) labels/classes. The oversampling method used is DBSCAN-SMOTE, combining DBSCAN and SMOTE helps address imbalanced data with significant outliers [15]. DBSCAN can identify and remove outliers before using SMOTE to balance the data. In the data pattern learning stage, a cost-sensitive learning approach is utilized.

The main goal is to optimize the model's performance by considering the costs associated with incorrect predictions, such as false positives and false negatives[16]. This model employs two base classifiers: Random Forest and Naïve Bayes. Reweighting instances involves adjusting the weights of instances based on the associated error costs, making the model more sensitive to costly errors. Class weights are used to penalize misclassifications of certain classes more than others. Finally, the decision threshold for predicting the positive class is adjusted to minimize the total cost of misclassifications. The final process involves evaluation using a confusion matrix, with measurement parameters including accuracy, precision, recall, F1-Measure, and Area under Curve (AUC). Based on the previous problem, the researcher has established a framework to address the issue. The research framework is outlined as follows in Fig. 2.

A. Data Acquisition and Split dataset

The thermogram images utilized in this study were sourced from IEEE Data port (https://ieee-dataport.org/open-access/plantar-thermogram-database-study-diabetic-foot-complications). This dataset includes thermal images (thermograms) of the plantar region, collected from 122 subjects diagnosed with diabetes (DM group) and 45 non-diabetic subjects (CG group). In total, the dataset comprises 334 images of both left and right plantar areas. The control group's thermogram images display a unique temperature distribution pattern (butterfly pattern), with higher temperatures in the middle of the foot and gradually decreasing temperatures in other areas. In contrast, the diabetic group exhibits relatively higher temperatures across the entire foot.

B. Image Enhancement (Contrast Limited Adaptive Histogram Equalization)

Image enhancement methods are essential for increasing the visibility and contrast of important features in thermographic images, particularly those of diabetic feet. Numerous approaches, such as contrast stretching, have been studied in the literature to contrast enhance these images and histogram equalization[17]. Contrast Limited Adaptive Histogram Equalization (CLAHE) Image enhancement focuses on improving local contrast, particularly in areas with uneven illumination. Unlike Histogram Equalization (HE), which processes the entire image, CLAHE works on small blocks known as tiles, and combines the results smoothly to prevent boundary artifacts [18]. The steps in CLAHE begin with (Divide the Image), where the image is split into small blocks of a specific size, followed by (Apply Histogram Equalization) applied to each block. Next is (Clip Histogram), where the histogram of each block is clipped to a certain value to prevent excessive contrast enhancement. The mathematical model of CLAHE on RGB images based on [18].

C. CNN feature extraction

In this study, feature extraction employs transfer learning using the Inception V3 architecture. Inception V3, developed by Google researchers, is a convolutional neural network designed to improve computational efficiency and effectiveness [14]. The Inception V3 mathematical model can be described using fundamental operations in convolutional neural networks, including convolution, pooling, normalization, and activation [11], show in (1-5).

- Convolution:

$$output[i,j,k] = \sum_{m,n} input[i+m, j+n].filter[m,n,k] \quad (1)$$

- The standard convolution operation involves sliding filters over the entire image to extract features
- Pooling:

$$output[i,j,k] = max_{m,n} input[i+m, j+n, k] \quad (2)$$

Pooling decreases spatial dimensions by selecting the maximum value within a defined window.

- Batch Normalization

$$\hat{x}^{(k)} = \frac{x^{(k)} - \mu^{(k)}}{\sqrt{\left(\sigma^{(k)}\right)^2 + \epsilon}} \quad (3)$$

$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)} \quad (4)$$

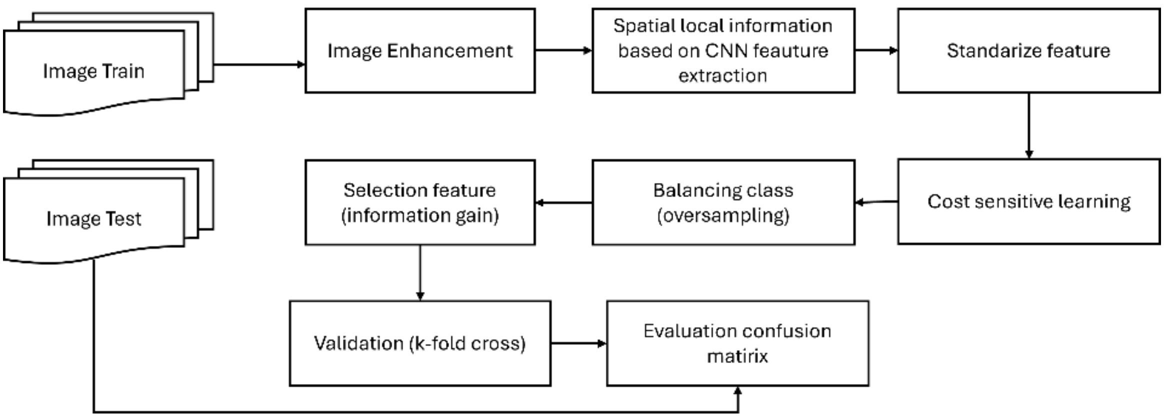

Fig. 2 Flow chart classification thermogram diabetic foot ulcer

where μ and σ is average and *batch variation*, γ and β is a learned parameter.

$$output = \max(0, input) \quad (5)$$

D. Selection Feature (information gain)

Information Gain is a filter-based feature selection method aimed at identifying the most relevant features for a specific class. This method helps reduce noise from irrelevant features by detecting those that provide the most information for a given class. The process involves determining the best attribute by first calculating its entropy, which measures class uncertainty using the probabilities of specific events or attributed. The information gain value for a feature is then calculated through several steps [20].

E. Balancing Class with SMOTE Oversampling

The Synthetic Minority Over-Sampling Technique (SMOTE) is a popular method used in machine learning to address the problem of imbalanced datasets. This imbalance can cause machine learning models to be biased toward the majority class, leading to poor performance on the minority class. SMOTE helps to mitigate this issue by generating synthetic examples of the minority class, show in (6).

Mathematical formulation of this methods is, Let (X_{minority}) be the set of all minority class samples, and (k) be the number of nearest neighbors:

- Select a minority class sample ($x_i \in X_{\text{minority}}$)
- Find (k) $-$ *nearest neighbors* ($N_k(x_i) \subset X_{\text{minority}}$)
- Randomly choose a neighbor ($x_{zi} \in N_k(x_i)$)
- Generate a synthetic example (x_{new}) as:

$$x_{\text{new}} = x_i + \delta \times (x_{zi} - x_i) \quad (6)$$

where ($\delta \sim \text{Uniform}(0,1)$)

F. Classification Algorithm

1) Cost Sensitive Random Forest: Cost-sensitive random forest algorithms (CSRFA) are designed to tackle imbalanced data categories and minimize misclassification costs by incorporating dynamic weights during feature selection and decision tree generation stages. These algorithms enhance the random forest model by embedding misclassification costs into the decision tree splitting index and refining optimization algorithms to achieve faster response times and higher classification accuracy, particularly in scenarios with unbalanced datasets[16]. The pseudocode for CSRFA showed in Table I.

2. Cost Sensitive Naïve Bayes: *Cost-sensitive Naive Bayes* (CSNB) is a variant of the Naive Bayes algorithm designed to account for varying misclassification costs. This algorithm integrates a cost matrix into the prediction process to minimize the total cost of misclassification [21]. The pseudocode for CSNB showed in Fig. 3.

G. Evaluation (k-fold cross validation)

K-Fold Cross-Validation is a statistical technique used to evaluate the performance of machine learning models. It helps in assessing how the model's results will generalize to an independent dataset, providing a robust mechanism to avoid problems such as overfitting and underfitting. K-Fold Cross-Validation divides the dataset into K distinct subsets (or "folds"). The model is then

***Pseudocode for* Cost Sensitive Random Forest**[22], [23]	**Pseudocode for Cost Sensitive Naïve Bayes** [21]
1. Initialize cost matrix C and determine instance weights W based on C. *2. For each tree in the forest:* *a. Perform weighted bootstrap sampling to create a training set.* *b. Build the decision tree:* *i. For each split in the tree:* *- Calculate the cost-sensitive splitting criteria using W and C.* *- Choose the split that minimizes the overall cost.* *ii. Handle missing values by assigning median values* *c. Train the decision tree on the weighted training set.* *3. Combine predictions from all trees using weighted voting.* *4. Evaluate the final model performance and adjust parameters as needed.*	*1. Preprocess data and determine the cost matrix C.* *2. Train the Naive Bayes model:* *a. Calculate prior probabilities P(C[k]).* *b. Calculate likelihood probabilities P(X[i] | C[k]) with smoothing.* *3. For each instance X in the test set:* *a. Calculate posterior probabilities P(C[k] | X) for each class.* *b. Calculate expected cost for each class:* *Expected Cost(C[k]) = sum(P(C[j] | X) * C[k[j]] for all j)* *c. Select the class with the lowest expected cost as the predicted class.* *4. Evaluate the model's performance.*

Fig. 3 Procedure for CSFRA and CSNB

trained and tested K times, each time with a different fold acting as the validation data and the remaining $K-1$ folds serving as the training data. Several classification evaluation metrics will be used to measure the model's performance [24], [25].

III. RESULT AND DISCUSSION

A. Experimental Settings

The experimental process is divided into three parts. The first part involves comparing the impact of oversampling to balance the data. The second part entails analysing the effect of feature selection. The final part involves comparing the results with several previous studies. Validation is performed using the k-fold cross-validation method with the number of folds set to 5. For model testing, the dataset is split into 70% training data and 30% testing data. The input images undergo Contrast Limited Adaptive Histogram Equalization (CLAHE) preprocessing. Fig. 4 showed the result of CLAHE.

B. Analysis on Different Oversampling Methods

The impact analysis of oversampling, aimed at data balancing, involves evaluating and comparing its effects on machine learning methods. The first scenario evaluates the model's performance without oversampling. The dataset comprises 90 images in the normal class (control group) and 247 images in the abnormal class (Diabetes Mellitus). The images are resized to 224 x 224 pixels, with a batch size of 32. The Adam optimizer is employed, and the loss function utilized is binary cross-entropy. Experimental results on the original dataset indicate an accuracy of 80.30% and an AUC of 82.5%. Fig. 4 illustrates the performance of the DFU classification model in the absence of data oversampling. Table II demonstrates the performance of the model obtained from the experiment without implementing data oversampling. The findings reveal that the model still exhibits misclassification errors between diabetic and nondiabetic data, achieving a precision of 83.0% for the nondiabetic class and 81.0% for the diabetic class. Furthermore, the recall values are consistent, with both the diabetic and nondiabetic classes recording a recall of 82.0%. These results highlight the need for further refinement to improve classification accuracy.

The second experiment investigates the effect of data oversampling by employing the original SMOTE method. The results demonstrate a notable enhancement in the performance of the cost-sensitive learning model due to the addition of synthetic data. Fig. 5 illustrates the accuracy and AUC values for this model. By introducing variability in the minority class data, SMOTE effectively mitigates overfitting, a common issue where the model becomes overly tailored to a small and overly specific subset of examples. This approach enhances the model's generalizability and robustness. The inclusion of synthetic data resulting from the oversampling process has shown a significant improvement in all performance metrics of the model. SMOTE generates synthetic examples for the minority class, providing the machine learning model with a larger dataset to learn from that class. This typically enhances the accuracy for the minority class, as the model becomes less biased towards the majority class. Specifically, there is an observed increase in recall performance by approximately 10% compared to the previous model. This demonstrates that the addition of synthetic data tends to enhance the sensitivity (recall) for the minority class, as the model is more likely to recognize examples from this class. However, a common consequence of using synthetic data is the increase in recall often accompanied by a decrease in precision. This occurs because more minority class examples are predicted, including incorrect predictions. This trade-off is important to consider when evaluating the overall effectiveness of synthetic data in improving model performance. In this second experiment, the accuracy achieved for the Nondiabetic

TABLE II
PERFORMANCE OF THE MODEL WITHOUT OVERSAMPLING

	Precision	Recall	F1-Measure	Accuracy
Non-Diabetic	0.830	0.820	0.850	0.830
Diabetic	0.810	0.820	0.810	0.810

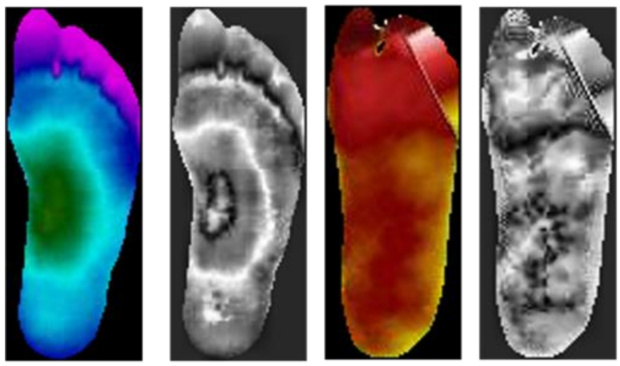

Fig. 4 The results of image preprocessing using the CLAHE method are as follows. (a) normal images (control group), (b) normal images after applying CLAHE, (c) abnormal images (diabetic mellitus), and (d) abnormal images after applying CLAHE

class was 98.1%, and for the Diabetic class, it was 96.1%, as shown in Table III. However, in the DFU dataset, this phenomenon did not occur. The precision values increased alongside the recall values, indicating that the use of synthetic data is effective for enhancing the performance of the cost-sensitive learning model on the DFU dataset. Subsequently, further experiments were conducted to explore various SMOTE development methods. The third experiment utilized the SMOTE Borderline method. SMOTE Borderline is an advanced variant of the SMOTE technique that aims to address class imbalance by focusing specifically on examples at the border between majority and minority classes. The core of the SMOTE Borderline method is to enhance the quality of synthetic data by considering the position of minority examples relative to majority examples, thereby producing more representative and effective samples for improving model performance. In this test, the distance metric used was Euclidean distance with k=5. Based on the test results, the synthetic data generated from SMOTE. Borderline yielded performance similar to that of the SMOTE method. In this third experiment, the accuracy achieved for the Nondiabetic class was 97.6%, and for the Diabetic class, it was 97.1%, as shown in Table IV. Additionally, as depicted in Fig. 6, there was no overfitting observed between the training and testing data. It can be concluded that the synthetic data from SMOTE Borderline has a comparable impact on the cost-sensitive learning classifier as the SMOTE method. Furthermore, the feature extraction used in this study was sufficiently effective in representing the characteristics of each class, resulting in minimal overlapping data at the boundary region between diabetic and nondiabetic classes. This finding underscores the effectiveness of SMOTE Borderline in enhancing model performance without compromising data integrity showed in Fig. 7.

TABLE III
PERFORMANCE OF THE MODEL WITH ADDITION OVERSAMPLING SMOTE

	Precision	Recall	F1-Measure	Accuracy
Non-Diabetic	0.960	0.970	0.991	0.981
Diabetic	0.970	0.970	0.970	0.961

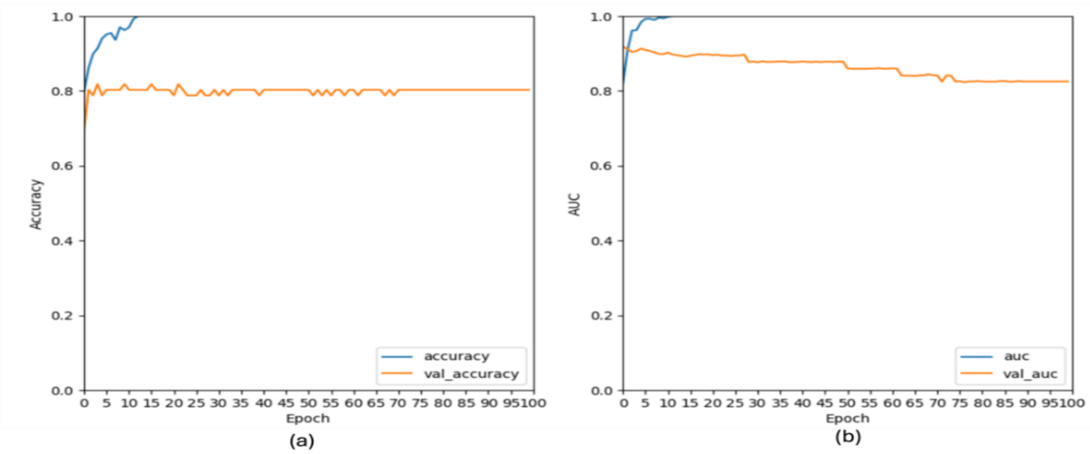

Fig. 5 Plot of accuracy (a) and AUC; (b) values against epochs for the model without data oversampling

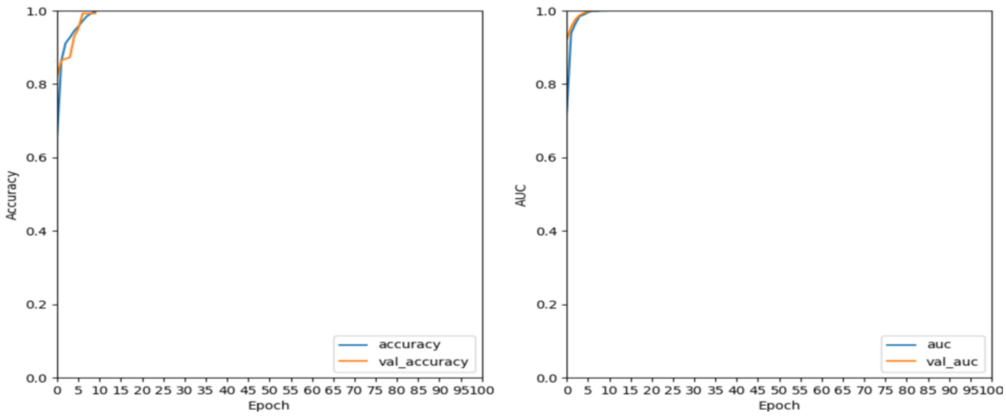

Fig. 6 Plot of accuracy (a) and AUC (b) values against epochs for the model with addition oversampling SMOTE

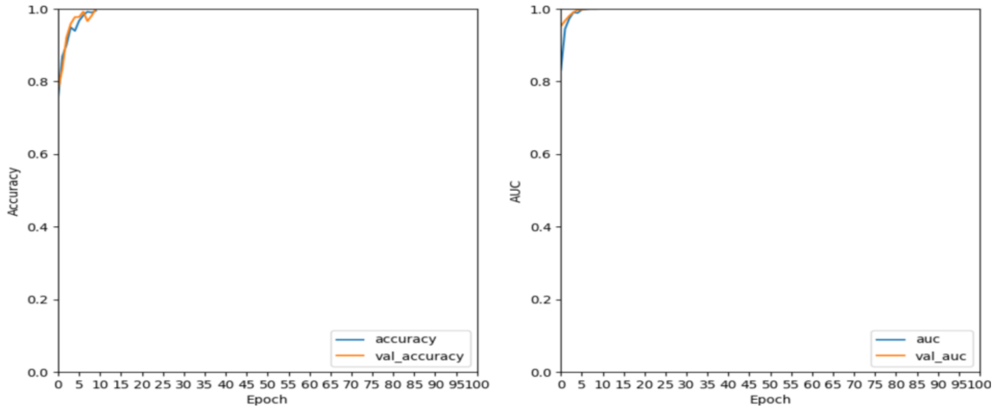

Fig. 7 Plot of accuracy (a) and AUC (b) values against epochs for the model with addition oversampling SMOTE borderline

TABLE IV
PERFORMANCE OF THE MODEL WITH ADDITION OVERSAMPLING SMOTE BORDERLINE

	Precision	Recall	F1-Measure	Accuracy
Non-Diabetic	0.970	0.972	0.970	0.976
Diabetic	0.981	0.972	0.971	0.971

The final experiment incorporated the SMOTE-IPF oversampling method. This approach aims to eliminate oversampled data that are unrepresentative for class pattern recognition. IPF (Iterative Partitioning Filter) is a technique used to remove less representative or noisy samples from the dataset, thereby enhancing the dataset's quality by ensuring only truly representative samples remain. Experimental results revealed that the integration of the IPF technique significantly improved the performance of cost-sensitive learning in recognizing data patterns from both diabetic and nondiabetic classes. The performance metrics of the cost-sensitive learning model in this experiment are presented in Table V. The model achieved an accuracy of 98.7%, with a precision of 99.1% for the diabetic class and 98.7% for the nondiabetic class. The recall values were 98.2% for the diabetic class and 98.1% for the nondiabetic class. The F-Measure was 98.1% for both classes.

Furthermore, there was no indication of overfitting, as evidenced by the model's comparable performance on training and testing data, as shown in Fig. 8. Although the improvement in performance was not markedly dramatic, the detection and removal of noisy data in the post-oversampling phase positively impacted the cost-sensitive learning model's ability to recognize patterns from both classes. These results demonstrate the efficacy of the SMOTE-IPF method in enhancing model performance by refining the quality of synthetic data, thus providing a more robust and accurate classification framework. The results from all the experiments conducted are summarized in Table VI and illustrated in Fig. 9. The data indicate that the combination of the Cost-Sensitive Model with the addition of SMOTE-IPF exhibited the best performance compared to the other models. On average, the performance improvement when comparing models without SMOTE to those with SMOTE was approximately 14-15%. This enhancement demonstrates that, for the DFU dataset, the application of the SMOTE oversampling technique not only increases the number of samples in the minority class but also ensures that the final dataset consists of more representative and high-quality samples. Consequently, machine learning models trained on this enriched dataset exhibit greater accuracy and reduced bias. These findings underscore the significance of employing advanced oversampling methods such as SMOTE-IPF to bolster the robustness and performance of machine learning models, particularly in contexts involving imbalanced datasets.

TABLE V
PERFORMANCE OF THE MODEL WITH ADDITION OVERSAMPLING SMOTE IPF

	Precision	Recall	F1-Measure	Accuracy
Non-Diabetic	0.987	0.981	0.981	0.987
Diabetic	0.991	0.982	0.981	0.987

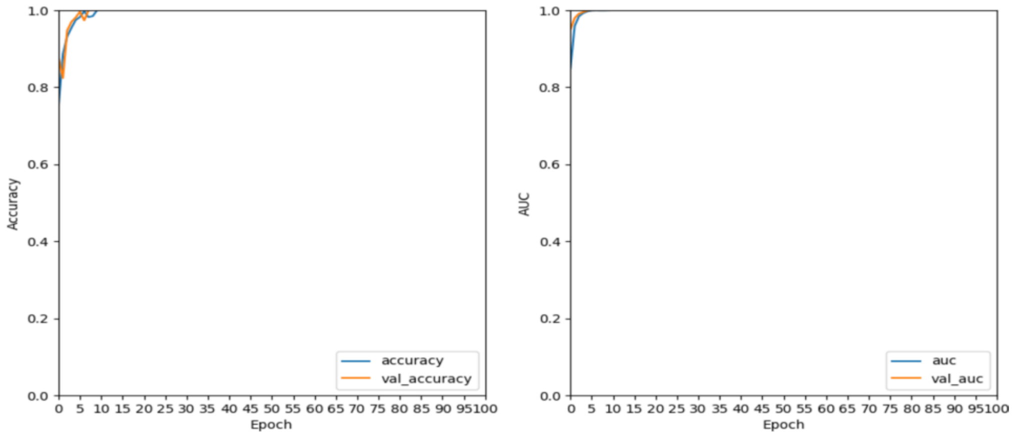

Fig. 8 Plot of accuracy (a) and auc (b) values against epochs for the model with addition oversampling SMOTE IPF

TABLE VI
COMPARISON OF COST-SENSITIVE LEARNING MODEL PERFORMANCE WITH VARIOUS SMOTE ENHANCEMENTS

	Diabetic				**Non-Diabetic**			
	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F-measure (%)**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F-measure (%)**
Non-Oversampling	0.830	0.830	0.820	0.850	0.810	0.810	0.820	0.810
Cost Sensitive Model + SMOTE	0.981	0.960	0.970	0.991	0.961	0.970	0.970	0.970
Cost Sensitive Model + SMOTE_Border	0.976	0.970	0.972	0.970	0.971	0.981	0.972	0.971
Cost Sensitive Model + SMOTE_IPF	**0.987**	**0.987**	**0.981**	**0.981**	**0.987**	**0.991**	**0.982**	**0.981**

The efficacy of SMOTE-IPF in improving classification performance underscores its potential as an invaluable tool in the preprocessing pipeline for developing reliable and generalizable predictive models. Next, we conducted a visual analysis of the distribution of synthetic data resulting from oversampling. This analysis compares the data distribution before oversampling, with the addition of SMOTE, SMOTE Borderline, and SMOTE-IPF. Fig. 10 presents scatter plots of the data distribution before and after oversampling. The scatter visualization utilizes features numbered 712 and 90 out of the total 2048 features used.

Based on the analysis, SMOTE-IPF is identified as the most effective technique for class separation while maintaining representative clusters, making it a superior choice for addressing imbalanced data. Although SMOTE enhances class representation, it does so to a lesser extent and results in greater class overlap. The original dataset, devoid of any oversampling techniques, exhibits significant class overlap, highlighting the necessity of methods such as SMOTE or SMOTE-IPF. This study underscores the critical importance of employing advanced oversampling techniques like SMOTE-IPF to enhance class separation and representation in imbalanced datasets. By creating a more balanced and representative dataset, these techniques significantly contribute to the development of more accurate and reliable machine learning models capable of effectively distinguishing between different classes. The enhanced class separation achieved through SMOTE-IPF ensures that the minority class is better represented, reducing bias and improving overall model performance. This finding reinforces the value of advanced oversampling methods in the preprocessing pipeline for imbalanced datasets, ultimately leading to more robust and generalizable predictive models.

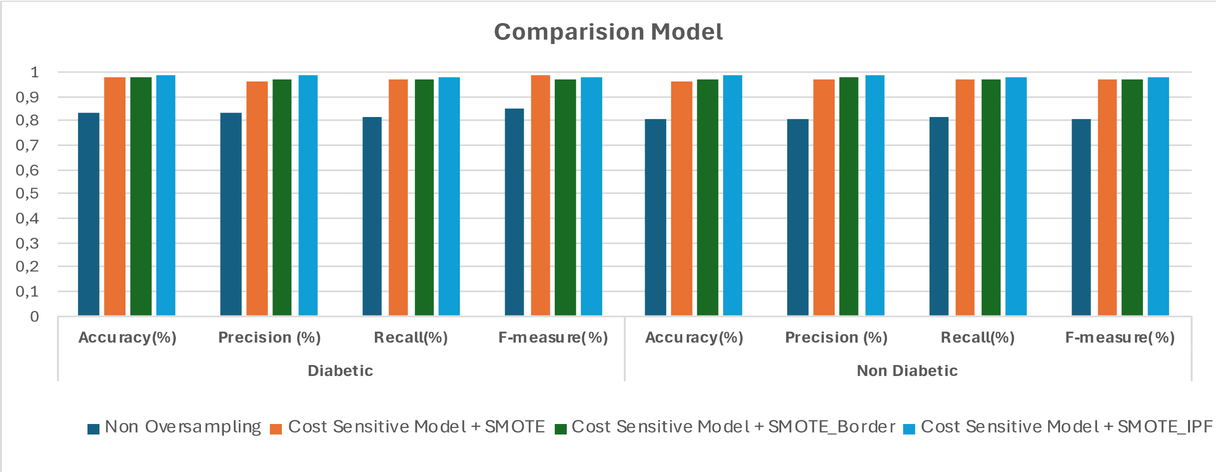

Fig. 9 Comparison graph of model performance with various SMOTE enhancements

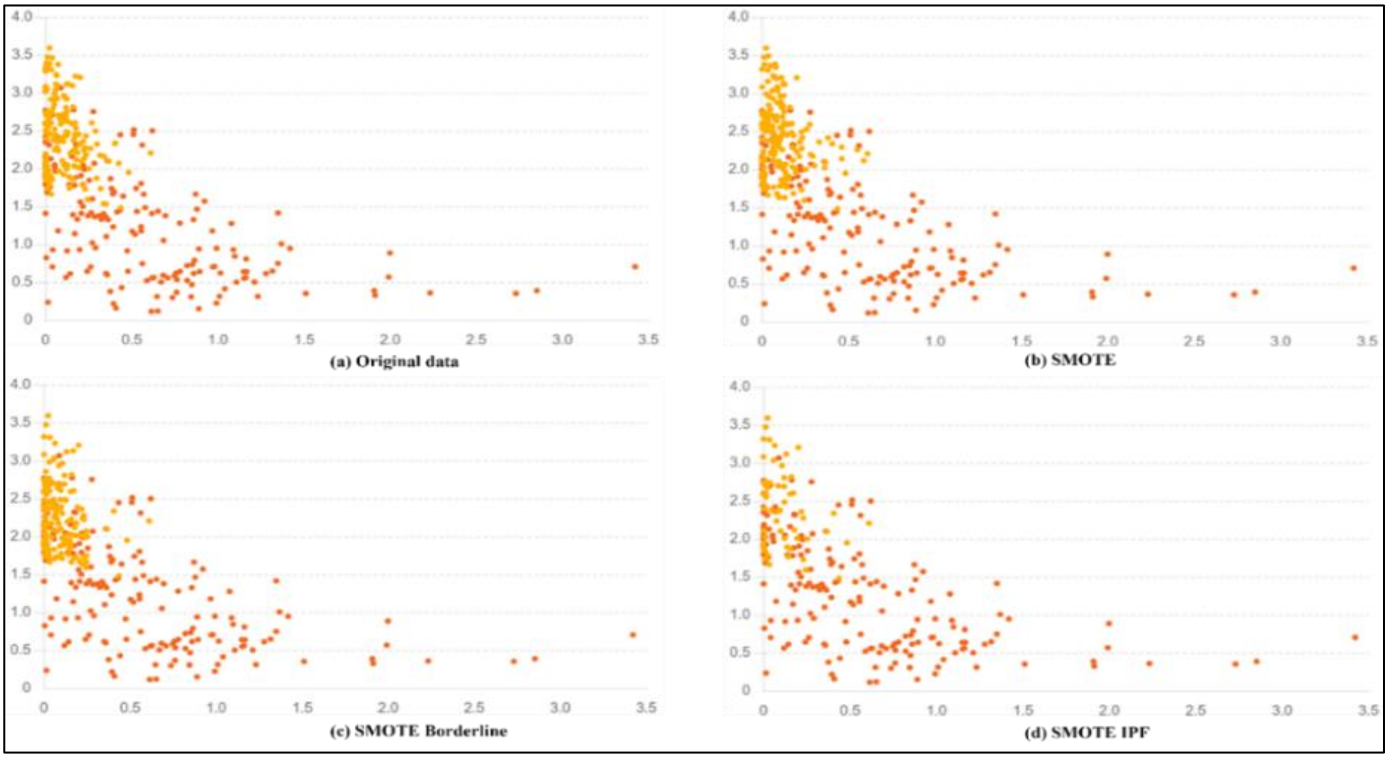

Fig. 10 Scatter plot visualization for data before and after oversampling

IV. CONCLUSION

This research explores the development and implementation of cost sensitive learning models for the classification of DFUs. Additionally, we address the challenges associated with class imbalance in medical datasets and evaluate the effectiveness of oversampling techniques such as SMOTE, SMOTE Borderline and SMOTE IPF. This study proposes a pretrained model method with InceptionV3 architecture to automatically obtain features in the DFU thermogram dataset. InceptionV3 as a feature extractor generally yielded satisfactory results, with an accuracy of 83.1% on non-diabetic data and 81.1% on diabetic data. In a subsequent experiment, incorporating the SMOTE oversampling technique led to improved performance, with accuracy rates rising to 98.1% for non-diabetic data and 96.1% for diabetic data. SMOTE IPF showed an accuracy of 98.7% was achieved, with precision rates of 99.1% for the diabetic class and 98.7% for the non-diabetic class. The recall rates were 98.2% for the diabetic class and 98.1%

for the non-diabetic class. The F-Measure was 98.1% for both the diabetic and non-diabetic classes. This enhancement demonstrates that applying the SMOTE oversampling technique to the DFU dataset not only increases the minority class samples but also ensures that the final dataset contains more representative and higher-quality samples. This allows the machine learning model trained on this dataset to be more accurate and less biased. The best results were obtained by combining cost-sensitive learning models with SMOTE IPF, showing that using the IPF method to reduce noise in the oversampled data successfully enhanced the model's performance

ACKNOWLEDGEMENT

The authors are grateful to the Directorate Jenderal Akademik of Research, Technology, and Community Service (DPRM) KEMENDIKBUDRISTEK Indonesia for funding this study through Program Funding (Penelitian Dosen Pemula) for the 2024 Fiscal Year.

REFERENCES

[1] J. Saminathan, M. Sasikala, V. B. Narayanamurthy, K. Rajesh, and R. Arvind, "Computer aided detection of diabetic foot ulcer using asymmetry analysis of texture and temperature features," *Infrared Phys Technol*, vol. 105, no. January, p. 103219, 2020, doi: 10.1016/j.infrared.2020.103219.

[2] D. Hernandez-Contreras, H. Peregrina-Barreto, J. Rangel-Magdaleno, J. Ramirez-Cortes, and F. Renero-Carrillo, "Automatic classification of thermal patterns in diabetic foot based on morphological pattern spectrum," *Infrared Phys Technol*, vol. 73, pp. 149–157, 2015, doi: 10.1016/j.infrared.2015.09.022.

[3] V. Filipe, P. Teixeira, and A. Teixeira, "Automatic Classification of Foot Thermograms Using Machine Learning Techniques," *Algorithms*, vol. 15, no. 7, 2022, doi: 10.3390/a15070236.

[4] M. Adam, NG E, Oh S, Heng M, "Automated characterization of diabetic foot using nonlinear features extracted from thermograms," *Infrared Phys Technol*, vol. 89, pp. 325–337, 2018, doi: 10.1016/j.infrared.2018.01.022.

[5] C. Evangeline N, S. Srinivasan, and E. Suresh, "Development of AI classification model for angiosome-wise interpretive substantiation of plantar feet thermal asymmetry in type 2 diabetic subjects using infrared thermograms," *J Therm Biol*, vol. 110, no. February, p. 103370, 2022, doi: 10.1016/j.jtherbio.2022.103370.

[6] M. H. Alshayeji, S. ChandraBhasi Sindhu, and S. Abed, "Early detection of diabetic foot ulcers from thermal images using the bag of features technique," *Biomed Signal Process Control*, vol. 79, no. P2, p. 104143, 2023, doi: 10.1016/j.bspc.2022.104143.

[7] J. J. Van Netten, M. Prijs, J. G. Van Baal, C. Liu, F. Van Der Heijden, and S. A. Bus, "Diagnostic values for skin temperature assessment to detect diabetes-related foot complications," *Diabetes Technol Ther*, vol. 16, no. 11, pp. 714–721, 2014, doi: 10.1089/dia.2014.0052.

[8] M. H. Yap, Haciuma R, Alavi A, "Deep learning in diabetic foot ulcers detection: A comprehensive evaluation," *Comput Biol Med*, vol. 135, p. 104596, 2021, doi: 10.1016/j.compbiomed.2021.104596.

[9] M. H. Yap, Classidy B, Byra M, "Diabetic foot ulcers segmentation challenge report: Benchmark and analysis," *Med Image Anal*, vol. 94, p. 103153, 2024, doi: 10.1016/j.media.2024.103153.

[10] P. N. Thotad, G. R. Bharamagoudar, and B. S. Anami, "Diabetic foot ulcer detection using deep learning approaches," *Sensors International*, vol. 4, no. October 2022, p. 100210, 2023, doi: 10.1016/j.sintl.2022.100210.

[11] M. H. Alshayeji, S. ChandraBhasi Sindhu, and S. Abed, "Early detection of diabetic foot ulcers from thermal images using the bag of features technique," *Biomed Signal Process Control*, vol. 79, no. P2, p. 104143, 2023, doi: 10.1016/j.bspc.2022.104143.

[12] P. D. W. Ayu, G. A. Pradipta, R. R. Huizen, E. S. W. Kadek, and I. G. E. Artana, "Combining CNN Feature Extractors and Oversampling Safe Level SMOTE to Enhance Amniotic Fluid Ultrasound Image Classification," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 1, pp. 251–262, 2024, doi: 10.22266/ijies2024.0229.24.

[13] G. A. Pradipta, R. Wardoyo, A. Musdholifah, and I. N. H. Sanjaya, "Radius-SMOTE: A New Oversampling Technique of Minority Samples Based on Radius Distance for Learning From Imbalanced Data," *IEEE Access*, vol. 9, pp. 74763–74777, 2021, doi: 10.1109/access.2021.3080316.

[14] A. S. Ali Abdullah Yaser Issam Aljanabi, "Developing a convolutional neural network for classifying tumor images using Inception v3." doi: 10.15587/1729-4061.2023.281227.

[15] N. L. Fitriyani, M. Syafrudin, G. Alfian, C. K. Yang, J. Rhee, and S. M. Ulyah, "Chronic Disease Prediction Model Using Integration of DBSCAN, SMOTE-ENN, and Random Forest," *2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems, ICETSIS 2022*, pp. 289–294, 2022, doi: 10.1109/ICETSIS55481.2022.9888806.

[16] J. M. Johnson and T. M. Khoshgoftaar, "Cost-Sensitive Ensemble Learning for Highly Imbalanced Classification," *Proceedings - 21st IEEE International Conference on Machine Learning and Applications, ICMLA 2022*, pp. 1427–1434, 2022, doi: 10.1109/ICMLA55696.2022.00225.

[17] U. S. Kumar, J. Simon, R. P. Vengaloor, and M. A. Elaveini, “Image Processing Techniques in Thermal and Non-thermal Images,” *Lecture Notes in Networks and Systems*, vol. 300 LNNS, no. January, pp. 533–544, 2022, doi: 10.1007/978-3-030-84760-9_45.

[18] G. Yadav, S. Maheshwari, and A. Agarwal, “Contrast limited adaptive histogram equalization based enhancement for real time video system,” *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014*, pp. 2392–2397, 2014, doi: 10.1109/ICACCI.2014.6968381.

[19] A. S. Ali Abdullah Yaser Issam Aljanabi, “Developing a convolutional neural network for classifying tumor images using Inception v3.” doi: 10.15587/1729-4061.2023.281227.

[20] M. Liang and W. Ahmad, “Breast Cancer Intelligent Diagnosis based on Subtractive Clustering Adaptive Neural Fuzzy Inference System and Information Gain,” *2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*, no. x, pp. 152–156, 2017.

[21] A. Ibáñez, C. Bielza, and P. Larrañaga, “Cost-sensitive selective naive Bayes classifiers for predicting the increase of the h-index for scientific journals,” *Neurocomputing*, vol. 135, pp. 42–52, 2014, doi: 10.1016/j.neucom.2013.08.042.

[22] Y. Chen, “Research on Cost-sensitive Classification Methods for Imbalanced Data,” *Proceedings - 2021 International Conference on Artificial Intelligence, Big Data and Algorithms, CAIBDA 2021*, pp. 224–228, 2021, doi: 10.1109/CAIBDA53561.2021.00054.

[23] D. Devi, S. K. Biswas, and B. Purkayastha, “A Cost-sensitive weighted Random Forest Technique for Credit Card Fraud Detection,” *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*, pp. 1–6, 2019, doi: 10.1109/ICCCNT45670.2019.8944885.

[24] P. D. W. Ayu and S. Hartati, “Pixel Classification Based on Local Gray Level Rectangle Window Sampling for Amniotic Fluid Segmentation,” *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 1, pp. 420–432, 2021, doi: 10.22266/IJIES2021.0228.39.

[25] D. S. N. Ayu, P D W, Sri hartati, Aina Musdholifah, “Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm,” in *Soft Computing in Data Science*, Springer Nature, 2022, p. 15. doi: https://doi.org/10.1007/978-981-16-7334-4_20.