AI-Enabled Infrared Thermography: Machine Learning Approaches in Detecting Peripheral Arterial Disease

Georgi Kostadinov[1,2](✉)

[1] New Bulgarian University, 21 Montevideo str, 1618 Sofia, Bulgaria
georgi.kostadinov@kelvin.health
[2] Kelvin Health, 47A Cherni Vrah blvd, 1407 Sofia, Bulgaria

Abstract. Peripheral Arterial Disease (PAD) is a common circulatory problem that, if undetected or untreated, can lead to severe health consequences, including amputation. This study presents a novel approach to PAD detection using thermal data collected via a mobile thermal camera, processed, and analysed through various machine learning algorithms. The investigation focused on six machine learning models: Linear Regression, Decision Trees, Random Forest, Neural Network, XGBoost, and LightGBM, and their ability to predict the presence of PAD based on thermal features extracted from different angiosomes of the legs. Each model was trained and validated on a dataset consisting of thermal data from 42 patients, annotated with PAD status based on angiography diagnostics. The performance of each model was evaluated using eight metrics, including accuracy, sensitivity, and specificity. The results indicate that ensemble methods, particularly XGBoost and LightGBM, outperformed the other models with an accuracy of 96.8%. This research demonstrates the potential of thermal imaging coupled with machine learning for the detection of PAD, offering a non-invasive, accessible, and cost-effective diagnostic tool.

Keywords: Peripheral Arterial Disease · Machine Learning · Thermal Imaging · Predictive Models · XGBoost · LightGBM

1 Introduction

1.1 Challenges and Opportunities

Early detection and diagnosis of Peripheral Arterial Disease (PAD) remains a significant challenge in the medical field. Traditional diagnostic methods, often relying on patient-reported symptoms, may not be present until the disease has significantly progressed. However, the inception of modern machine learning (ML) methodologies presents an exciting opportunity to transform PAD diagnostics. Infrared (IR) thermography, a non-invasive tool capable of detecting subtle skin temperature changes, shows promise in detecting PAD [1], but the volume of data generated can be overwhelming.

T. Zlateva and G. Tuparov (Eds.): CSECS 2023, LNICST 514, pp. 159–170, 2023.
https://doi.org/10.1007/978-3-031-44668-9_12

Machine learning comes to the rescue in managing and interpreting this vast data. It can analyse thermal images, identify abnormalities indicative of PAD, and condense the most critical findings into a summarized report, effectively creating a 'synopsis' of a patient's thermal data. This innovative integration of AI and ML with infrared thermography not only increases efficiency but also significantly improves diagnostic accuracy, presenting a new way to revolutionize the early detection of PAD.

1.2 Related Work

The utilization of thermal imaging as a tool for detecting various medical conditions has been extensively studied in recent years [2], with a focus on its potential in diagnosing Peripheral Arterial Disease (PAD) and Diabetic Foot Ulcers (DFUs), given the high rise of patients with these conditions post-COVID.

For instance, in the study by [3], infrared thermography was used to evaluate the severity, functional capacity, and quality of life in patients at high risk for PAD. The study demonstrates the potential of thermography as a non-invasive tool for assessing patients with PAD. Meanwhile, [4] proposed a Support Vector Classification model for PAD identification using features extracted from infrared thermography images.

Various researchers have explored the application of ML techniques in the early diagnosis of DFUs using thermogram images. In a study by [5], asymmetric analysis of temperature features was used for detecting diabetic foot ulcer. The study reported an impressive sensitivity and specificity of 96.5 and 92.41% respectively. Similarly, [6] proposed a machine learning-based model for classifying thermal distribution patterns in the feet of diabetic patients. Several other studies, such as [7, 8], and [9], have further explored the potential of machine learning techniques in interpreting thermogram images for the detection of DFUs.

These studies, however, mostly focus on the detection of foot ulcers in diabetic patients and do not specifically target the detection of PAD. Furthermore, most of the models proposed in these studies, such as Support Vector Machines, are simpler compared to the ensemble and neural network approaches evaluated in this research. This paper also presents a thorough comparison of these machine learning approaches in their ability to detect PAD from thermographic data as well as analysis of four thermal features and their relation to the performance of such models that will help guide further research in the field.

The rest of the paper is structured as follows. Section 2 discusses the proposed work with an overview of the algorithms and metrics used. Section 3 presents the collection, annotation, and processing of the thermal dataset, which forms the basis of this research. Section 4 is explaining the mechanics of the six machine learning (ML) methodologies selected in the presented work, with a specific emphasis on their application in detecting Peripheral Arterial Disease (PAD) through IR thermography data. In Sect. 5, the metrics for the evaluation of the selected methodologies, as well as the results achieved are presented. Finally, in Sect. 6, conclusions are drawn, and the future work is presented.

2 Proposed Work

This paper is an extension of the work presented in [1] by providing a comprehensive overview of the technology and methodologies incorporated within the Artificial Intelligence Supported Infrared Thermography (AISIT) system.

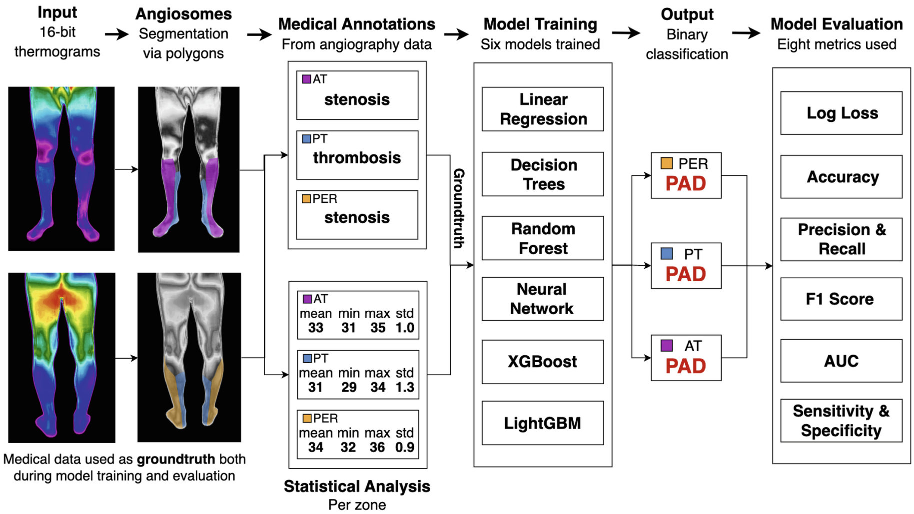

Fig. 1. Proposed work – six models were trained and evaluated by eight metrics in total.

The data acquisition process is detailed, illustrating how the infrared thermographic data is captured, processed, and prepared for further analysis by machine learning models. A discussion on which features are extracted from the thermal images is included.

Moreover, as visualised in **Fig. 1,** six various machine learning (ML) methodologies are explored, ranging from established techniques such as decision tree, random forest, as well as linear regression to more complex methods such as neural networks, XGBoost, and LightGBM. Each of these machine learning techniques is analysed and evaluated based on their ability to accurately identify PAD on IR data. This includes a comparative study detailing the performance metrics for each model, and an analysis on the feature importance used during training.

By offering a more in-depth exploration of the AISIT system from [1] and evaluating various machine learning methodologies, this paper aims to significantly contribute to the ongoing research in the field of PAD detection and diagnostics. Insights from this investigation can potentially drive the development of more accurate and efficient PAD detection systems based on AI and thermography.

3 Dataset

The study utilizes a dataset comprising thermal data from 42 patients, captured via a mobile thermal camera, FLIR One Pro. This dataset is based on the data used in the study in [1]. For each patient, thermal readings were taken from five different angles of the legs—front, back, left, right, and the feet. Additionally, the data collection process includes data from Angiography diagnostics.

The dataset from [1] contains a total of 313 medically and technically annotated angiosome regions of the legs of the 42 patients, with 78% of the regions identified as healthy and 22% - with pathology. Out of these 313 annotated samples, 60% or 188 randomly selected samples were used for training and the rest 125 - for validation purposes. The dataset encompasses a demographic variety [1], with 79% male patients, 21% females, all of Caucasian ethnicity, and a mix of additional factors, with 54% of the patients having diabetes and 76% being smokers.

The data collection process followed a strict medical protocol in a controlled hospital environment, as outlined in [1]. The thermal data was captured using the FLIR mobile thermal camera, taking readings from five different angles of the patient's legs. After that, each angiosome of the leg was manually annotated using the Supervisely[1] software, where polygonal segmentations were used to define each angiosome within the thermal image. Figure 1 demonstrates the process of collecting and annotating data for two angles of the legs – front and back, and three arterial regions, namely, for anterior tibial (AT), posterior tibial (PT), and peroneal (PER) arteries. The same process applies for all five angles and, in total, nine angiosomes of the legs. Each of the angiosome annotations of the legs corresponded with the results from the angiography diagnostics, which categorised each angiosome as healthy, stenosis, or thrombosis. For the purpose of this research, the categories of stenosis and thrombosis were combined into a single category labelled 'PAD,' leaving the healthy category as is.

The raw thermal images were processed into temperature readings using the raw2temp procedure from the R package Thermimage[2]. These temperature readings were subsequently used to extract thermal statistics from each segmented angiosome. Specifically, four thermal features were extracted: mean, minimum, maximum, and standard deviation of the temperature, as also shown on Fig. 1. These features form the basis for training the six ML models used in this study.

4 Machine Learning Methodologies

This section provides an overview of the theoretical foundations and practical implementation details of six machine learning methodologies. These models have been selected due to their ease of implementation, usability, and high interpretability. The emphasis on interpretability is of paramount importance given the context of medical AI systems, where the capability to provide explainable results is crucial. Thus, this section lays the groundwork for the subsequent analysis discussed in this research.

[1] Supervisely annotation software – https://supervisely.com.

[2] Thermimage R package - https://cran.r-project.org/web/packages/Thermimage.

4.1 Linear Regression

Linear regression is a fundamental method in statistics and machine learning [10], aiming to forecast a dependent variable based on a set of independent variables. The assumed linear relationship can be represented by the equation:

$$Y = \beta_0 + \beta_1 W_1 + \cdots + \beta_n W_n + \varepsilon \quad (1)$$

Here, Y is the dependent variable, W_1 through W_n are the independent variables, β_0 is the y-intercept, β_1 through β_n are the coefficients quantifying the influence of each W on W, and ε is the random error term. The implementation of the linear regression algorithm was done in the Python[3] programming language using the LinearRegression routines from the library Scikit-learn[4].

In the context of machine learning for Peripheral Arterial Disease (PAD) detection, linear regression can be used to determine the relationship between the four thermal data features (independent variables) and the presence or progression of PAD (dependent variable). The model is trained to minimize the differences between predicted and actual PAD status. Subsequently, it can predict PAD status based on new thermal data inputs.

4.2 Decision Tree

A decision tree [11] has a hierarchical model structure used in decision making and machine learning, consisting of nodes representing features, branches symbolizing decision rules, and leaves indicating outcomes. The root node, positioned at the top, partitions the data based on attribute values. This process runs throughout the tree, leading to a set of decision-making rules.

The decision tree algorithm used in this study is configured with several key parameters. The splitting criterion is set to Gini impurity, which measures the degree of class imbalance at each node. The maximum number of features used for splitting is set to 90% of the total features, while the minimum number of samples required to split a node is set at 30. The tree's maximum depth is limited to 4 to prevent overfitting. The log loss evaluation metric here is used to quantify the difference between predicted probabilities and true class labels. For the technical implementation of the algorithm, the DecisionTreeClassifier routines from the Scikit-learn library were used.

In the context of PAD detection with infrared thermography data, a decision tree can be trained to classify data into 'PAD' and 'healthy' groups based on the thermal features. The model partitions the data based on a series of binary decisions. The path from root to leaf provides a clear set of conditions leading to the prediction, making decision trees particularly interpretable, thus favourable for applications in the medical field.

4.3 Random Forest

A Random Forest [12] is an ensemble learning approach that is essentially a collection of decision trees, each constructed from a different subsample of the training data. The

[3] Python programming language - https://python.org.
[4] Scikit-learn Python package - https://scikit-learn.org.

final output of the model is derived by aggregating the predictions from all trees in the ensemble, typically using majority voting for classification tasks or averaging for tasks involving regression.

Using the same parameters as specified in 4.2, each decision tree within the Random Forest is built. The log loss evaluation metric, which quantifies the discrepancy between predicted probabilities and true class labels, is used for assessing the performance during training. The implementation was done using the RandomForestClassifier routines from the Scikit-learn library.

In the context of PAD detection, a Random Forest can utilize the variability in the thermal data features to train a multitude of decision trees. This ensemble approach provides a robust prediction model as it reduces the risk of overfitting seen in single decision trees and increases the generalizability of the model.

4.4 Neural Networks

Artificial neural networks (ANN) are hierarchical structures with layers consisting of nodes called neurons [13]. Each neuron applies a weighted sum to its inputs, the output of which is further calculated using a non-linear function, often referred to as the activation function.

In the presented work, a neural network with two dense, or fully connected, layers are utilized. The first dense layer contains 32 neurons, and the second contains 16. Each neuron in a dense layer is connected to every neuron in the previous layer, and the number of neurons in a layer can be seen as a measure of the layer's complexity or capacity.

The learning rate, a key parameter in training neural networks, is set at 0.05. This parameter determines the size of the steps taken during stochastic gradient descent [14] - the optimization algorithm commonly used to minimize the loss function in neural networks. The ANN architecture and training code was implemented using the PyTorch[5] library.

For the detection of PAD using infrared thermography data, the model is trained on the four thermal features in order to classify the data into the two categories. The log loss evaluation metric was used for minimizing the error rate. Despite its complexity, the neural network can provide a high level of accuracy in predicting PAD status, making it a valuable tool in this application.

4.5 XGBoost

The eXtreme Gradient Boosting (XGBoost) [15] model is a powerful ensemble learning method that leverages the concept of boosting weak learners. It is based on the gradient boosting framework but has been enhanced with a more regularized model formalization to control overfitting and increase performance.

In its essence, XGBoost constructs a strong predictive model, ensemble of multiple weaker predictive models, typically decision trees. Each new tree is grown to correct the residuals (the differences between the predicted and expected results) of the previous trees. The final prediction is calculated as the weighted sum of the predictions made

[5] PyTorch deep learning framework - https://pytorch.org.

by all the trees in the ensemble. The objective function of XGBoost that needs to be minimized can be represented as follows:

$$Objective = \sum_{i=1}^{N} \mathcal{L}\left(y_i, \hat{y_i}\right) + \sum_{k=1}^{K} \Omega(f_k) \tag{2}$$

where N are the number of data samples, K are the number of decision trees, $\mathcal{L}$ is the loss function that measures the difference between the actual and predicted values, and $\Omega(f_k)$ is the regularization term. The first term encourages the model to fit the data well, while the second term discourages overfitting. In the context of this research, $\mathcal{L}$ is the log loss detailed in the next section. The goal is to predict the probability of the instance being a positive class, which in this context refers to the presence of PAD.

The parameters of the XGBoost model have been tuned to optimize its performance on the specific task of detecting PAD using infrared thermography data. The learning rate is set to 0.075, controlling the shrinkage of the step size used in updates to prevent overfitting. The parameter for maximum depth of a tree is set to 6, balancing the model's complexity and its ability to learn fine-grained patterns. For the implementation of the algorithm, the Python library Xgboost[6] was selected for its ease of use and supported hardware.

XGBoost is a suitable choice for this study given its ability to handle a variety of data types, robustness to outliers, and capability to model complex non-linear relationships, all of which are critical in interpreting infrared thermography data for PAD detection. Its use of gradient boosting framework [15] also offers a robust mechanism for minimizing errors, offering high predictive accuracy.

4.6 LightGBM

LightGBM [16], standing for Light Gradient Boosting Machine, is another gradient boosting framework that employs tree-based methodologies. It is renowned for its speed and efficiency, as well as its suitability for handling large-scale data. LightGBM differs from other tree-based algorithms in its decision-making strategy, opting for a leaf-wise approach over the more conventional level-wise approach. This results in a more complex tree, but it also generally yields lower loss, contributing to better accuracy.

In the context of this project, the LightGBM model is configured for a binary classification task. The parameter for number of leaves, one of the critical parameters for model complexity, is set to 63. The learning rate is set to 0.05. The feature and bagging fraction parameters are both set to 0.9, meaning that 90% of the features and data are used at each iteration, respectively, adding an additional layer of randomness to make the model more robust to overfitting.

The minimum amount of data per leaf is set to 10. This parameter is a regularization measure that prevents the algorithm from creating leaves with fewer than 10 data points, further preventing overfitting. The training process is guided by the log loss metric, similarly to the models before. The implementation was done using the LightGBM[7] software library, maintained by Microsoft.

[6] Xgboost Python library - https://xgboost.readthedocs.io/

[7] LightGBM software toolkit - https://github.com/microsoft/LightGBM.

In summary, LightGBM's unique approach to tree building and flexible handling of different types of features, combined with its speed and efficiency, makes it an excellent tool for detecting PAD using infrared thermography data. The selected parameters are specifically aimed at reducing overfitting and achieving high predictive accuracy.

5 Results

5.1 Metrics

To evaluate and quantify the performance of each of the six models and facilitate comparison across them, eight metrics in total are employed. During the training phase, the log loss metric has been selected as the most appropriate for the binary classification task of detecting PAD. For validation purposes, the selected metrics are accuracy, precision, recall, F1 score, Area Under the Receiver Operating Characteristic curve (AUC), as well as sensitivity and specificity. Each of these metrics offers a unique perspective on the performance of the models, capturing different aspects of the prediction results.

Log Loss. The log loss [17] for binary classification problems can be computed as:

$$\mathcal{L} = -\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left(\mathrm{y_i}\log\left(\widehat{\mathrm{y_i}}\right) + (1-\mathrm{y_i})\log\left(1-\widehat{\mathrm{y_i}}\right)\right) \tag{3}$$

Here, N is the number of samples or instances, y_i is the actual class label (0 or 1 in binary classification) for label i, and $\widehat{y_i}$ is the predicted probability that the sample belongs to a particular class. Due to the logarithmic function, predictions that are close to the true values contribute a small amount to the overall loss, while predictions that are far off contribute significantly more.

Log loss provides a more nuanced view of the performance of a model than metrics like accuracy, as it considers the probability distributions of the predicted probabilities. Furthermore, comparing log loss to other loss metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE), it has the advantage of being more sensitive to the confidence of prediction [17], which can be a useful property in binary classification problems. However, it is also more sensitive to outliers and can be more difficult to interpret than some other metrics, like MAE or MSE.

Accuracy, Precision, Recall, F1 Score. These are the basic metrics commonly used in evaluating machine learning tasks [18]. Accuracy represents the ratio of correct predictions from the total number of predictions. While precision is the ratio of the true positive among all positive predictions, providing a measure of the ability to correctly detect cases with PAD, recall is the ratio of the true positive among all actual positive cases, reflecting the model's capacity to detect PAD cases. Finally, the F1 score is the harmonic mean of both precision and recall, making it a unifying metric that balances the two. It offers a more balanced measure when the class distribution is uneven.

AUC. Area under the ROC Curve (AUC) [19] is a robust metric for binary classification problems as it measures the ability to distinguish between the classes at various threshold settings. The ROC curve plots the true positive rates against the false positive rates at multiple thresholds, and AUC measures the entire two-dimensional area underneath this curve. An AUC of 1 signifies a perfect classifier, whereas an AUC of 0.5 speaks for a model equivalent to random guessing.

Sensitivity and Specificity. In the medical field, these are the two most common metrics that evaluate the performance of a binary classification test [20].

Sensitivity measures the ratio of actual positive cases that are correctly identified as such. It is extremely important in the medical field as it gauges the ability to correctly diagnose patients with a particular disease. The higher the sensitivity, the lower the chances of a false negative result. Sensitivity is calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

Specificity measures the ratio of actual negative cases that are correctly identified. In the medical context, this means the ability to correctly diagnose healthy patients as healthy. High specificity correlates to less false positives, which is important in order to avoid unnecessary treatments or interventions. Specificity is calculated as follows:

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

In the context of Peripheral Arterial Disease detection, high sensitivity ensures that patients with the disease are identified, reducing the risk of false negatives. This is particularly important in medical conditions where early detection can significantly improve the prognosis. On the other hand, high specificity ensures that healthy individuals are not falsely identified as having the disease, thus avoiding unnecessary further tests or treatments.

5.2 Feature Importance

For each model, a feature importance matrix was generated, revealing important insights about the significance of the four temperature-based features in predicting the presence of Peripheral Arterial Disease. As seen in Fig. 2, the standard deviation of temperature

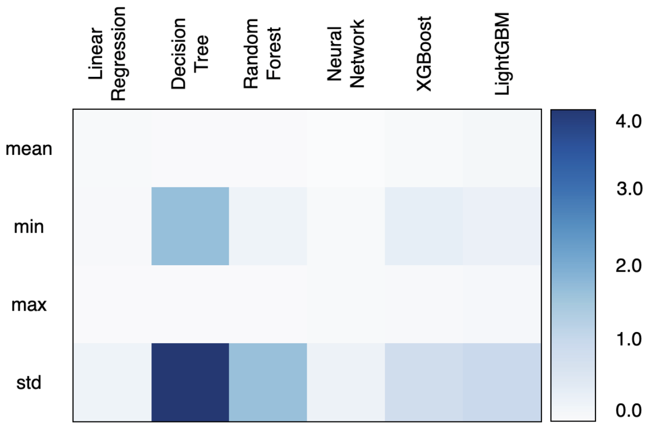

Fig. 2. Feature importance matrix for each model.

consistently emerged as the most influential feature across all machine learning models. This suggests that variation in temperature, rather than just average or extreme values, plays a crucial role in identifying PAD. The minimum temperature was typically the second most important feature, suggesting that the coldest temperature readings may be indicative of compromised blood flow, a characteristic symptom of PAD.

Interestingly, the mean temperature and maximum temperature demonstrated varying levels of importance across different models. While these features held some predictive power in models such as Linear Regression, XGBoost, and LightGBM, they were deemed unimportant by the Decision Trees and Random Forest models. This discrepancy underlines the complexity of PAD detection and the need for a multi-faceted approach.

Overall, these results, underscore the importance of considering the entire range of temperature characteristics - including the mean, extremes, and variability - in the thermographic analysis for PAD detection.

5.3 Performance Analysis

Table 1. Comparative model performance analysis for metrics log loss (Loss), accuracy (Acc.), precision (Pre.), recall (Rec.), F1 score (F1), AUC, sensitivity (Sens.), and specificity (Spec.).

Model	Loss	Acc	Pre	Rec	F1	AUC	Sens	Spec
Linear Regression	0.341	0.873	0.75	1.0	0.588	0.87	0.3	0.981
Decision Trees	0.267	0.873	1.0	1.0	0.512	0.857	0.2	1.0
Random Forest	0.2	0.92	0.667	1.0	0.8	0.977	1.0	0.906
Neural Network	0.358	0.857	0.6	1.0	0.533	0.825	0.3	0.962
XGBoost	0.141	0.968	1.0	1.0	0.909	0.975	1.0	0.962
LightGBM	0.111	0.968	1.0	1.0	0.909	0.985	1.0	0.962

The results of the ML methodologies trained on the thermal data for Peripheral Arterial Disease (PAD) detection are presented in Table 1. Each model was evaluated on all eight metrics. The results obtained from the trained models indicate varying degrees of performance. The performance metrics reveal different aspects of each model's ability to correctly predict the presence or progression of PAD.

Among the models, LightGBM and XGBoost stand out with the lowest log loss and the highest accuracy, precision, recall, F1 score, sensitivity, and specificity. This superior performance can be attributed to their gradient boosting mechanisms, which optimizes the models by iteratively adding weak learners, thereby reducing the bias and variance. Moreover, these models handle feature interactions well and are resistant to overfitting, which makes them particularly suitable for this dataset.

In contrast, Linear Regression and Decision Trees show fewer promising results. The relatively simple structure of these models may struggle with the complex patterns in the thermal data. While Decision Trees have perfect precision, its lower sensitivity indicates its difficulty in detecting true positive PAD cases.

The Random Forest model, an ensemble of Decision Trees, significantly improves the performance over a single Decision Tree, as shown by its lower log loss and higher accuracy, F1 score, and AUC. This improvement underscores the power of ensemble methods in handling complex datasets.

The Neural Network also demonstrates a reasonable performance, although it doesn't outperform LightGBM and XGBoost. This could be due to the relatively simple architecture that was chosen and its hyperparameters or the need for more data to effectively train the network.

These results have several implications for the field. First, they demonstrate the potential of machine learning, particularly advanced gradient boosting models, in detecting PAD using thermal data. This application could significantly enhance early detection and treatment of PAD, thereby improving patient outcomes. Second, these results may guide future research on PAD detection on thermal data, suggesting that focus could be directed towards optimizing models like LightGBM and XGBoost or exploring other complex models, such as deeper neural networks.

6 Conclusions

6.1 Summary

This study presented a comprehensive analysis of six ML models applied to a unique dataset of thermal images from [1] to detect Peripheral Arterial Disease (PAD). The dataset, based on thermal data from 42 patients, was processed and medically annotated, extracting four thermal features from each angiosome region for the training of the models. Importance analysis for each feature was also discussed, that would navigate future research on the topic.

Six models, namely Linear Regression, Decision Trees, Random Forest, Artificial Neural Networks, XGBoost, and LightGBM were trained and evaluated. A variety of metrics, including log loss, accuracy, precision, recall, F1 score, AUC, sensitivity, and specificity were employed to provide a holistic view of the models' performance. The results showed that ensemble models such as XGBoost and LightGBM generally outperformed other models, with LightGBM exhibiting the best overall performance.

6.2 Future Work

While the results of this research are promising, there is always room for growth and further exploration in the field. Future work could include the inclusion of more diverse data, encompassing patients from different ethnic backgrounds, varying age groups, and other lifestyle factors, to increase the generalizability of the models. The exploration of additional thermal features from each angiosome could be further looked into in order to improve the accuracy of the detection models. While six machine learning models were examined in this study, there are numerous other ML models that can be tested, such as convolutional neural networks (CNN), which have shown to perform significantly better in image-based tasks [21]. Lastly, future research could investigate the integration of these models into clinical decision support systems, providing a tool that can aid clinicians in diagnosing PAD. The findings of this study not only contribute to the growing body of research in the field of medically applied thermography and machine learning but also pave the way for future research in this area.

References

1. Stankov, Z., et al.: Infrared thermography—diagnostic method in peripheral artery disease. Arab J. Intervent. Radiol. **7**(S 01), A1217 (2023)
2. Wang, Q., et al.: Infrared thermography for measuring elevated body temperature: clinical accuracy, calibration, and evaluation. Sensors **22**(1) (2022)
3. Huang, C., et al.: The application of infrared thermography in evaluation of patients at high risk for lower extremity peripheral arterial disease. J. Vasc. Surg. **54**(4), 1074–1080 (2011)
4. Padierna, L.C., et al.: Classification method of peripheral arterial disease in patients with type 2 diabetes mellitus by infrared thermography and machine learning. Infrared Phys. Technol. **111**, 103531 (2020)
5. Saminathan, J., et al.: Computer aided detection of diabetic foot ulcer using asymmetry analysis of texture and temperature features. Infrared Phys. Technol. **105**, 103219 (2020)
6. Evangeline, N.C., et al.: Application of non-contact thermography as a screening modality for Diabetic Foot Syndrome – A real time cross sectional research outcome. Biomed. Signal Process. Control **79**, 104054 (2023)
7. Srinivasan, S., et al.: 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 **110**, 103370 (2022)
8. Khandakar, A., et al.: A machine learning model for early detection of diabetic foot using thermogram images. Comput. Biol. Med. **137**, 104838 (2021)
9. Maldonado, H., et al.: Automatic detection of risk zones in diabetic foot soles by processing thermographic images taken in an uncontrolled environment. Infrared Phys. Technol. **105**, 103187 (2020)
10. Seber, G.A.F., Lee, A.J.: Linear Regression Analysis, vol. 330. Wiley (2003)
11. Swain, P.H., Hauska, H.: The decision tree classifier: design and potential. IEEE Trans. Geosci. Electron. **15**(3), 142–147 (1977)
12. Breiman, L., et al.: Random forests. Mach. Learn. **45**, 5–32 (2001)
13. Kostadinov, G.: Synopsis of video files using neural networks. In: Proceedings of the 23rd EANN 2022 (2022). https://doi.org/10.1007/978-3-031-08223-8_16
14. Bottou, L.: Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22–27, 2010 Keynote, Invited and Contributed Papers. Physica-Verlag HD (2010)
15. Chen, T., et al: Xgboost: extreme gradient boosting. R Package Version 0.4–2 **1(4),** 1–4 (2015)
16. Ke, G., et al: LightGBM: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. **30** (2017)
17. Vovk, V.: The fundamental nature of the log loss function. Fields of Logic and Computation II: Essays Dedicated to Yuri Gurevich on the Occasion of His 75th Birthday, pp. 307–318 (2015)
18. Handelman, G.S., et al.: Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. Am. J. Roentgenol **212**(1), 38–43 (2019)
19. Hanley, J.A., McNeil, B.J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology **143**(1), 29–36 (1982)
20. Sharm, D., Yada, U.B., Sharma, P.: The concept of sensitivity and specificity in relation to two types of errors and its application in medical research. J. Reliabil. Stat. Stud. 53–58 (2009)
21. Li, Z., et al.: A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Trans. Neural Netw. Learn. Syst. (2021)