IEICE TRANSACTIONS

on Information and Systems

Machine Learning-based System for Heat-Resistant Analysis of Car Lamp Design

Hyebong CHOI[†a], *Member*, Joel SHIN[†], Jeongho KIM[†], Samuel YOON[†], Hyeonmin PARK[†], Hyejin CHO[††],
and Jiyoung JUNG[†††], *Nonmembers*

SUMMARY   The design of automobile lamps requires accurate estimation of heat distribution to prevent overheating and deformation of the product. Traditional heat resistant analysis using Computational Fluid Dynamics (CFD) is time-consuming and requires expertise in thermofluid mechanics, making real-time temperature analysis less accessible to lamp designers.
      We propose a machine learning-based temperature prediction system for automobile lamp design. We trained our machine learning models using CFD results of various lamp designs, providing lamp designers real-time Heat-Resistant Analysis. Comprehensive tests on real lamp products demonstrate that our prediction model accurately estimates heat distribution comparable to CFD analysis within a minute. Our system visualizes the estimated heat distribution of car lamp design supporting quick decision-making by lamp designer. It is expected to shorten the product design process, improving the market competitiveness.
key words: *automobile lamp, heat-resistant analysis, temperature prediction, machine learning*

1. Introduction

Heat-resistant analysis is a critical step in designing an automobile lamp. It helps determine whether the proposed design will be able to withstand the high temperatures generated by the lamp's bulb. By conducting this analysis, designers can identify potential issues and make necessary adjustments to ensure the final product meets the desired specifications and performs as intended. Conducting heat-resistant analysis prior to creating a prototype or product can save time and money by identifying potential issues and allowing for modifications to be made to the design before it goes into production.

      In industry, heat-resistant analysis is primarily performed based on computational fluid dynamics (CFD), which utilizes the Navier-Stokes equations, nonlinear partial differential equations that describe fluid motion with viscosity [1]. CFD uses numerical methods such as finite difference, finite volume, or finite element methods to discretize and solve the Navier-Stokes equations, which describe the motion of viscous fluids. The solution of these equations allows for the analysis of fluid flow and heat transfer problems in various engineering applications [2], [3], [4]. Generating 3D input grids for analysis, performing CFD-based heat-resistant analysis, and verification using specialized software require the assistance of thermal-fluid experts and can take several days. This process will be repeated as many as necessary to validate various design elements and can cause significant delays in the design process of car lamps.

      Machine learning is a technique used to develop intelligent systems that find patterns and make predictions by combining data and statistical modeling. Typically, a large amount of data is used to train a model, which is then used to make predictions on unseen datasets. The prediction model can be trained with numerous CFD analysis results for various car lamp designs, producing accurate estimations of heat distribution comparable to CFD analysis. This approach enables designers to verify their own designs in real-time without the assistance of an expert, improving the efficiency of the design process.

      In this study, we propose a machine learning-based system for predicting the thermal performance of automotive lamps. The proposed system first learns from a large amount of CFD analysis results for various lamp designs. The trained prediction model is then used to provide real-time temperature analysis results for new lamp designs and design parameters. The system provides an interface for designers to easily input design parameters such as bulb Wattage and item materials without the need for expert knowledge of heat transfer theory or specialized CFD programs. This eliminates the need for complex processes such as re-generating a 3D grid for analysis programs where the prediction model directly operates on CATIA models provided by the designer. Through this approach, designers can perform cost-effective evaluations of various design parameters by performing machine learning-based thermal analysis in the pre-review stage for CFD-based analysis. We have also used downsampling techniques to address the data imbalance problem where the majority of the data points in the training data are in the low-temperature range, while the high-temperature range, which is more critical for heat-resistant analysis, has relatively lower number of data points. This improved the scalability of the system, and we have verified the accuracy of the temperature prediction within a 7% error range compared to the results of CFD analysis.

      [†]The author is with School of Global Entrepreneurship and ICT, Handong Global University, Republic of Korea
      [††]The author is with School of Department of Industrial Engineering, Ulsan National Institute of Science Technology, Republic of Korea
      [†††]The author is with SL corporation, Republic of Korea
      a) E-mail: hbchoi@handong.edu

2. Related Work

2.1 Thermal Resistance Analysis in Car Lamp Design

Computational fluid dynamics (CFD) calculates mass conservation equations, momentum equations, and energy equations related to fluid flow using computers through numerical discretization and various algorithms [1]. CFD is a numerical model that approximates phenomena such as fluid flow, heat transfer, and related chemical reactions through computer-based simulation [5]. Heat treatment analyses utilizing CFD take into account crucial factors present in the actual heat treatment process, including phenomena like boiling, to ensure accurate computational results [6], [7], [8].

Thermal resistance analysis in car lamp design process primarily relies on Navier-Stokes equations in CFD, which are nonlinear partial differential equations that describe the flow of viscous fluids. CFD methods are known for their high accuracy in temperature estimation and their ability to examine a wide range of design specifications [6].

However, it requires a considerable amount of time to perform CFD-based thermal analyses in the design process. Experts typically need several days to generate a three-dimensional input grid, execute and validate the simulations, and perform iterative analyses to verify different design elements. Lamp designers often encounter challenges in performing the analysis independently, as it requires assistance from thermal fluid theory and analysis program experts. In the industry, reducing the time spent on the product development process is a pivotal factor in sustaining business competitiveness. Therefore, reducing the time required for thermal analyses in lamp design can make a substantial impact.

Hai Guo [9] proposed a deep learning-based temperature prediction model to prevent aging or damage in Permanent Magnet Synchronous Motors (PMSM). PH Gunawan, D Munandar [10] utilized Long Short-Term Memory (LSTM) to predict air temperature changes in Indonesia, considering its diverse climate conditions. Lee [11] developed an algorithm that computes heat distribution in car lamp design using complex physical formulas. However, due to computational complexity, the analysis has been performed on a simplified box-shaped lamp design rather than real lamp design.

We propose a machine learning-based temperature prediction system for thermal analyses as an alternative to CFD. This system enables designers to independently verify various design factors and obtain visualized real-time analysis results.

2.2 Ensemble Algorithm

The ensemble method is a learning algorithm that combines a set of weak learners to achieve an accurate prediction model by reducing bias and variance of the model [12]. It collects results from multiple predictors and integrates the result to produce the final answer.

There are three main ensemble learning methods: bagging, boosting and stacking. Bagging, short for Bootstrap Aggregating, is an ensemble method that improves the accuracy and stability of predictive models [13]. It generates multiple subsets of the training data through random sampling with replacement, trains separate models on these subsets, and combines their predictions to make the final prediction. By training models on different subsets of the data and aggregating their predictions, bagging helps to reduce overfitting, increase robustness, and improve generalization. Random Forest is the most well-known algorithm where bagging is applied to decision tree model [14].

Boosting aims to improve the performance of weak or base models by combining them into a strong predictive model. Unlike bagging, which trains models independently, boosting trains models sequentially. Each subsequent model focuses on the instances that were misclassified by previous models. In this way, boosting puts more emphasis on difficult-to-predict instances, allowing the models to learn from their mistakes and improve over iterations. The final prediction is made by aggregating the predictions of all the models. Boosting algorithms include Adaboost[15], Catboost[16], XGBoost[17], and LightGBM[18]. They have shown capability in capturing complex patterns and improve prediction accuracy.

Stacking, also known as stacked generalization, combines the predictions of multiple individual models to obtain a more accurate and robust model [19]. Unlike traditional ensemble methods that rely on simple averaging or voting, stacking involves training a meta-model that learns to combine the predictions of the base models. The base models are trained on the original data, while the meta-model is trained on the predictions made by the base models. Stacking leverages the strengths of different models and can capture complex relationships in the data, resulting in improved predictive performance [20].

In this study, we used the CV(cross-validation)-based stacking ensemble algorithm to estimate the temperature and heat distribution of car lamp design. This improves the accuracy of estimation and avoids overfitting problem.

2.3 Data Re-sampling

Data re-sampling plays a crucial role in addressing imbalanced datasets, where one class or a range of values is significantly underrepresented compared to the others. Training with imbalanced datasets can lead to biased model performance in supervised learning. Re-sampling techniques aim to balance the data distribution and improve model performance.

Oversampling is a commonly used technique in data resampling. It involves increasing the number of instances in the minority class to match the number of instances in the majority class. ROSE and SMOTE are well-known oversampling methods. Random Oversampling(ROSE) method is a sampling technique for solving binary classification data

imbalance problems. ROSE provides a unified framework to address model estimation and accuracy issues simultaneously, with the advantage of reducing the risk of overfitting.[21]

The Synthetic Minority Oversampling Technique (SMOTE) is a statistical method that generates synthetic examples by interpolating between neighboring instances of the minority class [22]. SMOTE is applied when analyzing classes with insufficient representative values [23]. It is less likely to cause overfitting compared to random oversampling and helps to preserve information without reducing the dataset size, unlike undersampling.

Downsampling, on the other hand, is a technique used to address data imbalance by extracting meaningful data from a myriad of data. The advantages of downsampling are that it improves computational efficiency by reducing the size of the dataset, reduces storage space by reducing the amount of data, and reduces noise in the data.[24].

In thermal resistance analysis, we attempt to estimate a continuous variable, temperature of each data point. There are significantly larger portion of data points in a low temperature range whereas less than 5% of data points are near to or exceeding the temperature limit of the lamp materials. To address this imbalance and focus on the crucial high temperature range, we applied downsampling to the training dataset. The random reduction of low temperature range data points allows for greater emphasis on the high temperature range, which is critical for assessing the lamp design's ability to withstand heat. It also drastically cuts down the computation cost and time in training models.

3. Methodology

3.1 Data Description

This study primarily focuses on utilizing machine learning to create a predictive model for estimation of heat distribution of automobile lamp designs. To train this model, Computational Fluid Dynamics (CFD) temperature analysis results obtained from actual lamp designs of 11 real car models serve as the primary data source. For the reliability of the predictive model, we have generated various analysis cases altering design parameters such as voltage of bulbs and material color for each car lamp design. The total number of analysis cases created is 526 from the 11 unique car models for model training.

We trained the predictive model mainly on the CFD analysis result of rear lamp design. The rear lamp design is further divided into two categories: single bulb and double bulb configurations. We restricted the scope of our prediction models within these two types of configurations for simplicity even though there are more design options that contain more than 2 bulbs in a lamp design. Depending on its function, a single bulb can be lighted or two bulbs can be turned on simultaneously in a thermal resistance analysis of lamp design.

	Single Lamp	Double Lamp
Housing	25,846	27,017
Reflector	17,215	19,636
Inner Lens	15,184	13,099
Outer Lens	12592	13,757
Heat Plate	2,317	3,674

Table 1 Average number of data points for each lamp item

A rear lamp design consists of 5 parts: housing frame, inner lens, outer lens, reflector, and heat plate. Table 1 shows the average number of data points of each part in single and double lamp configuration. The data points represent temperature measured at intervals of 3 mm in the Cartesian coordinate system. On average, the housing frame, the largest part of lamp contains the highest number of coordinate data points. The heat plate, the smaller part, has the fewest data points.

	Variable	Description	Variable	Description
Spec	Ambient Temperature of Housing	mean: 53.66 sd: 19.75 min: 25 max: 85	Ambient Temperature of Lens	mean: 53.66 sd: 19.75 min: 25 max: 85
	Function	STOP(298 / 56%) TAIL(10 / 2%) T/STOP(152 / 30%) T/SIG(10 / 2%) B/UP(54 / 10%)	HDT	mean: 132.93 sd: 28.8 min: 91 max: 360
Bulb	On Time	mean: 22.04 sd: 26.65 min: 5 max: 60	Off Time	mean: 3.02 sd: 7.84 min: 0 max: 55
	Wattage	mean: 23.5 sd: 11.98 min: 5 max: 38	Bulb Type	Single Bulb(207 / 40%) Double Bulb(317 / 60%)
	Bulb Center Coordinate	3D Position Center Coordinates of the Bulb in the Lamp	Lighting Conditions	Continuous(398 / 76%) Discontinuous(126 / 24%)
Lamp	Item	Housing(524 / 25%) Reflector(524 / 25%) Inner Lens(410 / 20%) Lens(524 / 25%) Heat Plate(91 / 5%)	Raw Material Color	White(3 / 0.1%) Gray(267 / 13%) Yellow(410 / 20%) Red(632 / 30%) Black(806 / 40%) Clear(91 / 4%)
	External Surface Absorption Coefficient	mean: 0.37 sd: 0.34 min: 0.06 max: 0.89	Internal Surface Absorption Coefficient	mean: 0.37 sd: 0.34 min: 0.08 max: 0.9
	Heat Plate	Not Exist(323 / 62%) Exist(201 / 38%)	Hole	Not Exist(234 / 45%) Exist(290 / 55%)
	LED	Not Exist(122 / 23%) Exist(402 / 77%)		

Table 2 Dataset description

In thermal resistance analysis in car lamp design, there are three types of factors that possibly affect the heat distribution. The factors consist of experiment specification(Spec), bulb characteristics(Bulb), and lamp design factors(Lamp) as illustrated in Table 2

Firstly, the experiment specification is the test condition required by client that includes ambient temperature and heat deflection temperature (HDT). Those factors specify the experimental condition that the actual lamp products are to pass. Variables illustrate the experimental environment setting and the function, as well as the heat deflection temperature (HDT) of the lamp. "Ambient_temp_hsg" and "ambient_temp_lens" denote the ambient temperature surrounding the housing frame and lens of the lamp. The

"Function" variable represents the lamp's functions that can be used in different driving situation including "stop", "turn signal", and "tail". "HDT" implies the temperature limit that lamp material can endure, in other words, the heat deflection temperature.

As the second type of factors are related to the lamp bulbs. They are fundamental characteristics of the bulb within a lamp. This includes bulb center position (in x, y, and z coordinate), wattage, bulb type (single or double), bulb lighting conditions (continuous or non-continuous), and on/off interval.

Variable	Description
X	data point position in x-coordinate
Y	data point position in y-coordinate
Z	data point position in z-coordinate
Temperature (Celsius degree)	Temperature derived from CFD Analysis mean: 86.274 sd: 37.04 max: 359.854 min: 4.407

Table 3 Coordinate Dataset description

The wattage variable is crucial as it determines the brightness of the lamp as well as amount of heat generated from the bulb. The bulb center position is typically the hottest point in the lamp where heat is transferred throughout the lamp.

Lastly, there are lamp design factors. The variable "item" implies the components of the lamp including housing, reflector, inner lens, outer lens, and heat plate. "Heat plate" is a binary variable indicating the presence of a heat plate, which functions to dissipate the generated heat. Variables "Hole" and "LED" indicate the presence of a heat release hole or LED module. The designer may add a heat-release hole and a heat plate to reduce the excessive heat. The "Raw_material_color" variable is the color of the material for the lamp item which determines its absorption coefficient meaning how much the item surface absolves the heat. The "Surface Absorption Coefficient" variable, categorized as the Interior/Exterior Surface Absorption Coefficient, is the absorption coefficient of the surface.

With CFD-based analysis, temperature of lamp items are estimated at every 3 mm interval. Those estimated temperatures are collected as training dataset described in Table 3. This temperature-coordinate dataset contains CFD-analyzed temperatures "temp_c" and their corresponding x,y, and z coordinates. The coordinate dataset is combined with the dataset of lamp design specification in Table 2. For double lamp configuration, two different bulbs generate simultaneously light and heat in the lamp. As illustrated in Figure 1, variables related to the two different bulbs are put together and joined with the temperature-coordinate dataset.

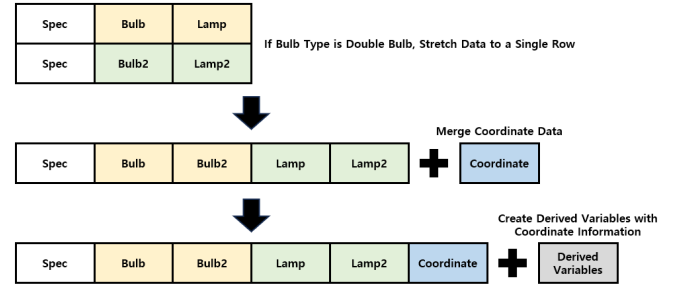Furthermore, we have extracted additional variables to

Fig. 1 Data Transformation for Double Lamp Configuration

improve the prediction model. We have calculated the Euclidean distance between the coordinate and the bulb center position, which can significantly affect the heat transfer. Equation 1 calculates the Euclidean distance from bulb center position to each data points where x_i, y_i, z_i represent x, y, and z coordinates of each data point and x_c, y_c, z_c are the coordinates of the bulb center position. x, y, and z coordinate distances are also introduced to include directional effect of heat transfer as calculated in Equation 2 where d_i^x, d_i^y, d_i^z are distance in each coordinate.

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2} \qquad (1)$$

$$\begin{aligned} d_i^x &= x_i - x_c \\ d_i^y &= y_i - y_c \\ d_i^z &= z_i - z_c \end{aligned} \qquad (2)$$

The volume of lamp design is also an important factor in determining heat distribution. However, it is too costly to compute the exact volume of the irregular shape of lamp design. We simplified the volume approximation by calculating it as the product of the x, y, and z coordinate ranges for the lamp design, as shown in Equation 3.

$$\begin{aligned} \text{Approx. Volume} &= (\max(x_i) - \min(x_i)) \times \\ &(\max(y_i) - \min(y_i)) \times (\max(z_i) - \min(z_i)) \end{aligned} \qquad (3)$$

3.2 Temperature Prediction Model

3.2.1 CV-based Stacking Ensemble

For accurate estimation of heat distribution, we have employed Cross-Validation (CV) based stacking ensemble algorithm [25]. The algorithm represents an enhanced version of the conventional stacking ensemble method, specifically designed to address the overfitting problem often encountered with the traditional approach. The conventional stacking ensemble method utilizes the complete prediction values for modeling, which can be prone to overfitting, especially with complex or high-dimensional datasets. To tackle the
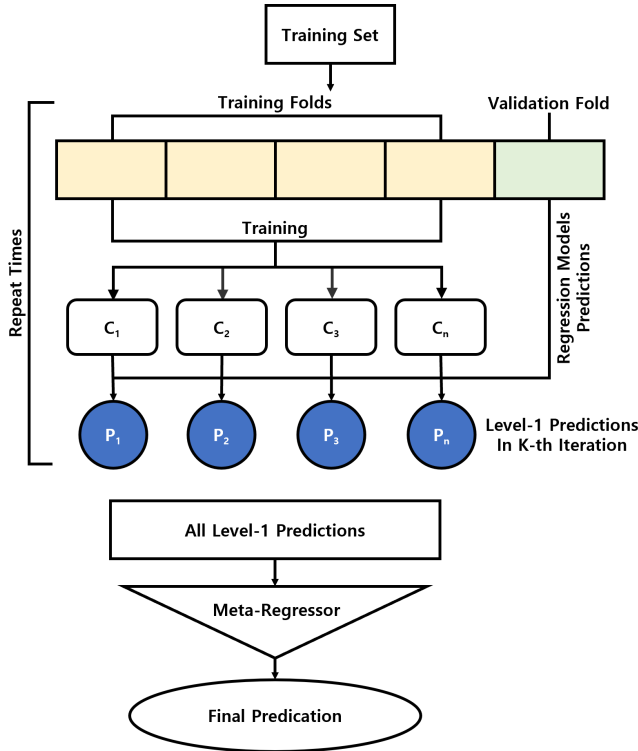
Fig. 2 CV-based Stacking Ensemble Overview

problem, the CV-based stacking ensemble algorithm was suggested. This method begins by dividing the training data into k-folds, where k represents the number of car cases in this particular context. Each fold alternates in serving as a validation set, while the remaining folds comprise the training set. Contrary to the traditional method, this cyclic process enables predictions to be made across the entire training dataset via the CV method, generating Meta Train Data. This approach mitigates overfitting and improves the overall performance of the stacking ensemble method.

Figure 2 illustrates the process of the CV based stacking ensemble algorithm. It shows how the training set is segmented into distinct training and validation folds. Upon this division, multiple models, denoted as $C_1, C_2, ..., C_n$ where n stands for the number of models, are trained on these folds. Prediction results, $P_1, P_2, ..., P_n$, are derived from each model C_i across all car cases. These prediction results collectively form the meta train data, which is less prone to overfitting as compared to the predictions obtained from the traditional stacking ensemble method. In the subsequent phase, the test data is fed into the k learning models trained from k-folded training data, yielding k-dimensional data. To align its format with the meta train data, an average is taken for the k-dimensional data. This process is repeated for n multiple models, resulting in "stacked" meta train data and meta test data with m columns each. If there are m models, the resulting "stacked" meta train data and meta test data will each incorporate m columns. In the final stage, the meta train data is trained to create a meta learner or the final model.

Final predictions are then generated from the meta test data.

3.2.2 Max Temperature Prediction Model

Although the CV-based Stacking Ensemble algorithm produces accurate estimation of heat distribution on overall dataset, the training process is still affected by imbalanced dataset where the data points in the range of the temperature near to or over the limit (HDT) are relatively scarce than those of lower temperature range. Imbalanced training dataset leads the model to be more optimized for learning in the low-temperature range, which may result in less accurate estimation for the part in high-temperature range that is crucial in heat resistance analysis.

To compensate this limitation, we separately train a model that makes prediction on the maximum temperature of each part of lamp design using XGBoost algorithm. We provide the estimated maximum temperature to the designer as well as to the CV-based Stacking Ensemble model for the estimation of the heat distribution as an input variable to improve the prediction accuracy on the high temperature range. By incorporating the predicted maximum temperature into the CV-based Stacking Ensemble model, we were able to generate more accurate and reliable temperature predictions for thermal resistance analysis.
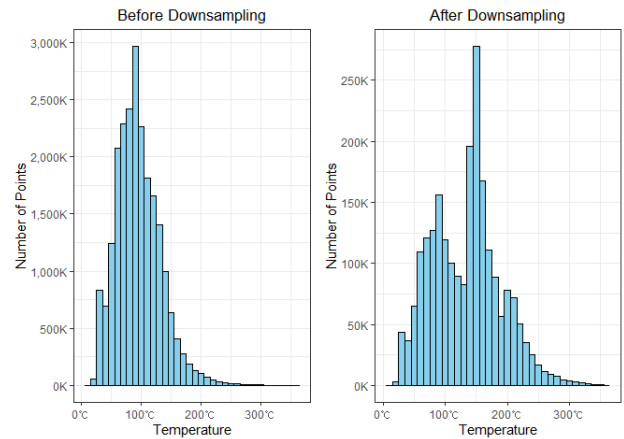
3.2.3 Downsampling

Fig. 3 Distribution of Lamp Temperature Before and After Applying Downsampling

We have down-sampled the training dataset to mitigate the data imbalance problem where the majority of data points are concentrated in the low-temperature range, while high-temperature data is sparse. The training dataset is divided into two partitions based on temperature: the top 5% representing high-temperature data and the bottom 95% representing low-temperature data. Subsequently, we randomly sampled 5% of data points from the bottom 95% and concatenate them with the top 5%, resulting in a downsized training dataset that is 1/10 the size of the original.

As depicted in Figure 3, down-sampling process adjusts the distribution of data points' temperature range, leading the CV-based Stacking Ensemble model to be more optimized for high temperature range. This adaptation makes the model more suitable for thermal resistance analysis, where it is critical to identify which parts of the lamp design may exceed the temperature limit. Furthermore, it significantly reduces the training time while enhancing the prediction accuracy of the high-temperature region without compromising the overall accuracy of heat distribution estimation. In the next section, we demonstrate that applying down-sampling method effectively improve the performance of heat resistance analysis while considerably saving the model training time.

3.3 System Architecture
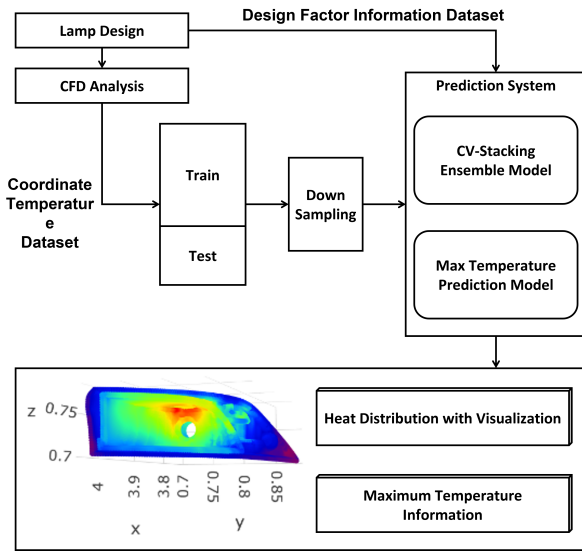
Fig. 4 System Architectures

Figure 4 shows the architecture of the machine learning-based heat resistant analysis system. It consists of two main processes: one that trains temperature prediction model and the other one that perform thermal resistance analysis for lamp design.

To extract the dataset for training the temperature prediction model, we conducted CFD analysis on various real car lamp designs. Then, we partitioned the dataset into two: one for training model and the other for validation. To address data imbalance, down-sampling was applied to the training dataset. The training dataset was used to learn the max temperature prediction model and the CV-based Stacking Ensemble model, which estimate heat distribution. These models were then validated using the test dataset.

When utilizing the system for thermal resistance analysis, designers are required to input various design factors, such as bulb voltage and materials, along with a CATIA 3D mesh into the system. The system then generates predictions for the maximum temperature and heat distribution of the lamp design. These results are presented through an interactive 3D visualization created using the R Shiny package,

allowing designers to assess whether their design meets the thermal resistance requirements. Additionally, the system provides real-time analysis results, enabling designers to dynamically adjust lamp design factors, modify their design, and evaluate its performance.

4. Experiments

4.1 Experiment Setting

We have trained and verified the prediction model in the system environment specified in Table 4. We mainly used R version 4.2.2 with its machine learning libraries.

	System Environment
OS	Windows 11 Pro 21H2 Version
CPU	12th Gen Intel(R) Core(TM) i9-12900
GPU	NVIDIA GeForce RTX 3090 X 2EA
RAM	Samsumg 32GB DDR4 25600 X 4EA
R	4.2.2 Version
Packages	h2o 3.34.0.7, lightgbm 3.2.1, xgboost 1.4.1, catboost 1.0.0, randomForest 4.7-1.1, glmnet 4.1-8

Table 4 Experiment environment

As an alternative to CFD-based thermal resistance analysis, the machine learning-based prediction models are developed to provide accurate estimations of heat distribution. We conducted a comparison of two stacking ensemble prediction models with five single models: LightGBM, Xgboost, Catboost, RandomForest and Least Absolute Shrinkage and Selection Operator (Lasso) [26]. The aim was to prove that the CV-based stacking ensemble models outperform other algorithms and are well-suited to heat prediction in car lamp design.

The models under consideration were: (A) another stacking ensemble model using nine XGB and nine LightGBM (LGBM) base models with an ANN as the final model, (B) a stacking ensemble model using 18 XGBoost (XGB) base models and an artificial neural network (ANN) as the final model, (C) a single Catboost model, (D) a single RandomForest model, (E) a single LightGBM model, (F) a single XGBoost model, and (G) a single Lasso model. The parameters of each model were tuned using grid search to optimize their performance.

For the model evaluation, we have used to the Mean Absolute Percentage Error (MAPE) metric that compares the differences between CFD-computed temperature and model-estimated temperature of data points. The formula to calculate MAPE is as follows:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{A_i - F_i}{A_i} \right| \tag{4}$$

where F_i is the model-predicted temperature of the car lamp at data point i, A_i is the CFD-computed temperature of the car lamp at data point i, and n is the number of data points that consist of lamp design shape.

In thermal analysis of car lamp design, it is crucial to ensure that the high-temperature regions do not exceed the heat resistance limit of the lamp material. To evaluate the prediction accuracy in the high-temperature regions, we measured the prediction errors of the top 5% highest temperature points and calculated the "hMAPE" (high MAPE). The hMAPE is computed as the Mean Absolute Percentage Error(MAPE) for the data points corresponding to the top 5% highest temperatures.

K-fold cross-validation was employed to evaluate the performance of the predictive model, where *k* represents the number of CFD analysis cases for car lamp design. Each lamp design's CFD result was treated as a separate fold, with one case designated as the test data and the remaining cases as the training data. This method prevents data leakage and bias when examining the model's generalization across various vehicle models. We used the Mean Absolute Percentage Error as the performance metric, both for all data points and specifically for the data points within top 5% of highest temperature range. The metric explains how closely the model's predicted temperatures are aligned with the results of the CFD analysis.

4.2 Performance Analysis

	MAPE	hMAPE
S.E.D XGB/LGBM	**6.90%**	**5.40%**
S.E.D XGB	7.02%	8.03%
CatBoost	7.30%	8.27%
Randomfoest	8.41%	8.41%
LightGBM	9.34%	8.4%
XGBoost	8.85%	10.15%
Lasso	25.45%	13.65%

Table 5   Performance comparison in single-bulb setting

	MAPE	hMAPE
S.E.D XGB/LGBM	**4.25%**	**2.51%**
S.E.D XGB	4.70%	3.39%
CatBoost	6.31%	4.60%
Randomfoest	5.41%	8.95%
LightGBM	7.7%	4.96%
XGBoost	7.06%	4.65%
Lasso	19.15%	10.99%

Table 6   Performance comparison in double-bulb setting

We trained and validated separate prediction models for two types of rear lamp designs: single-bulb and double-bulb configurations. Table 5 and Table 6 present a comparison of temperature prediction accuracy for different models under each configuration. These tables show the performance of all data points in the lamp design and the data points within the top 5% temperature range. In the tables, the model abbreviations represent the following:

- S.E.D (XGB/LGBM): CV-based stacking ensemble algorithm using nine XGBoost (XGB) and nine Light-GBM (LGBM) base models with an ANN as the final model

- S.E.D (XGB): CV-based stacking ensemble model using 18 XGBoost (XGB) base models and an artificial neural network (ANN) as the final model
- CatBoost: CatBoost used as a single prediction model
- RF: RandomForest used as a single prediction model
- LGBM: LightGBM used as a single prediction model
- XGB: XGBoost used as a single prediction model
- LASSO: Lasso used as a single prediction model

For both the single-bulb and double-bulb configurations, the Cross-Validation (CV) based stacking ensemble models, specifically S.E.D (XGB) and S.E.D (XGB/LGBM), demonstrate superior performance over the individual prediction models for the entire dataset. Notably, the ensemble model that integrates XGBoost with LightGBM (S.E.D (XGB/LGBM)) achieves the highest prediction accuracy within the critical high-temperature range, indicating its effectiveness and robustness for thermal resistance characterization. The prediction accuracy in high temperature range is more crucial for the lamp designer to evaluate their design whether it can resist the thermal limit of the material.

	Single Lamp	Double Lamp
S.E.D (XGB/LGBM)	3.107 sec	1.500 sec
S.E.D (XGB)	3.685 sec	2.082 sec
CatBoost	0.226 sec	0.933 sec
Randomfoest	1.975 sec	2.837 sec
LightGBM	0.009 sec	0.011 sec
XGBoost	0.0845 sec	0.462 sec
Lasso	0.834 sec	1.283 sec

Table 7   Average time spent for estimating heat distribution

In conclusion, the stacking ensemble algorithm using XGBoost and LightGBM delivers the most accurate estimation of heat distribution, particularly in the high temperature range. This enables designers to perform robust and reliable heat resistance analysis. The stacking ensemble models, incorporating maximum temperature models and 18 different types of boosting models, outperform single models, effectively addressing data imbalance and mitigating the risk of overfitting, especially in the high-temperature region. As shown in Table 7, even for complex models such as CV-based stacking ensemble, the average time spent on estimating heat distribution is within a few seconds, enabling real-time thermal resistance analysis.

Down-sampling was applied to the training dataset to emphasize data points in the high temperature range during model training. The impact of the down-sampling process is demonstrated in Table 8 and Table 9. Given the impractical training times without down-sampling, we conducted a comparison by using a downsized training dataset that was 1/20th the size of the original dataset. The tables compare the prediction accuracy with MAPE for overall datapoints and time spent for training models with original and downsampled training dataset. They indicate a notable improvement in prediction accuracy as a result of down-sampling, along with a significant reduction in training time by 4 to

and stress relief heat treatment of an aero engine component, Finit Elements Engrg. Des., 2003.

[8] H.Alberg, Material modeling for simulation of heat treatment, Department of Applied Physics and Mechanical Engineering, LULEÅ University of Technology, 2003.

[9] H. Guo, Q. Ding, Y. Song, H. Tang, L. Wang, and J. Zhao, Predicting Temperature of Permanent Magnet Synchronous Motor Based on Deep Neural Network, Energies, 2020.

[10] P. Gunawan, D. Munandar, and A. Farabib, Long Short-Term Memory Approach for Predicting Air Temperature In Indonesia, 2020.

[11] J. Lee, D. Choi, Y. Kim, S. Choi, Y. Park, J. Lee, and K. Lee, Development of Temperature Prediction Simulation Program for Automotive Headlamp, Korea Automotive Engineers Spring Conference, 2020.

[12] T.G. Dietterich, Ensemble Methods in Machine Learning, International Workshop on Multiple Classifier Systems, 2000.

[13] L. Breiman, Bagging predictors, Machine Learning, 1996.

[14] L. Breiman, "Random forests," Machine learning, vol.45, pp.5–32, 2001.

[15] Y. Freund and R.E. Schapire, XGBoost: A Scalable Tree Boosting System, journal of computer and system sciences, 1996.

[16] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, and A. Gulin, CatBoost: unbiased boosting with categorical features, NeurIPS, 2018.

[17] T. Chen and C. Guestrin, XGBoost: A Scalable Tree Boosting System, ISBN, 2016.

[18] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.Y. Liu, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, NIPS, 2017.

[19] D.H. Wolpert, "Stacked generalization," Neural networks, vol.5, no.2, pp.241–259, 1992.

[20] I. Syarif, E. Zaluska, A. Prugel-Bennett, and G. Wills, Application of bagging, boosting and stacking to intrusion detection, International Workshop on Machine Learning and Data Mining in Pattern Recognition, 1996.

[21] G. Menardi and N. Torelli, Training and assessing classification rules with imbalanced data, Data Min Knowl Disc, 2012.

[22] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, SMOTE: Synthetic Minority Over-sampling Technique, Journal of Artificial Intelligence Research, 2002.

[23] L. Torgo, R.P. Ribeiro, B. Pfahringer, and P. Branco, SMOTE for Regression, Portuguese Conference on Artificial Intelligence, 2013.

[24] W. Wei, X.L. Yang, B. Zhou, J. Feng, and P.Y. Shen, Combined Energy Minimization for Image Reconstruction from Few Views, Mathematical Problems in Engineering, 2012.

[25] S.A. Tang, Jiliang and H. Liu, Data Classification: Algorithms and Applications., CRC Press, 2014.

[26] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society Series B-Methodological, vol.58, pp.267–288, 1996.

Hyebong Choi is a Professor in the Global Entrepreneurship and ICT at Handong Global University. His research interests include scalable data mining, machine learning, decision support system, and industrial data analysis.

Joel Shin is an undergruate student in the Global Entrepreneurship and ICT at Handong Global University. His research interests lie at the crossroads of Industrial Engineering, AI in Ergonomics, and Tech Convergence in Social Analytics. His work aims to fuse artificial intelligence with ergonomic principles to increase industrial efficiency and safety. Furthermore, he explores how technological convergence can be leveraged in social analytics, promoting an interdisciplinary approach to better comprehend and optimize societal processes.

Jeongho Kim is currently a student at Handong Global University where he is majoring in Information Communication Technology (ICT) Convergence and Management. His research interests lie in the field of Industrial AI and MLops, focusing particularly on streamlining the gap between industries and AI technology. He seeks to leverage artificial intelligence to enhance the efficiency and productivity of various industrial processes.

Samuel Yoon is currently a student at Handong University, majoring in ICT convergence and statistics. His research interests are operational research and data mining for industrial process optimization.

Hyeonmin Park currently a student at Handong Global University, where he is majoring in Information Communication Technology (ICT) Convergence and Management. His research interests lie in the field of Startup Valuation, focusing on predicting and analyzing the startup ecosystem to provide valuable information for potential investors.

Hyejin Cho is a graduate student of Department of Industrial Engineering, Ulsan National Institute of Science & Technology. She received her B.A in Management B.E in ICT Convergence from Handong Global University. Her research interests are data mining for quality control and data-driven business strategy.

Jiyoung Jung is an engineer in thermal resistant analysis team in SL cooperation. His research interests include automation of lamp design and assessment process using machine learning and data collection process.