ABSTRACT The facial air temperature has a significant impact on the driver’s thermal comfort. Machine Learning models have been proved to be evidently effective in temperature predicting. In this study, three models are employed to predict the drivers’ facial temperature in a certain series of vehicles, which are Support Vector Regression (SVR), Artificial Neural Network (ANN), and Gated Recurrent Unit (GRU) respectively. We conduct an electric vehicle air-conditioning system experiment to collect the datasets of drivers’ head temperature and 6 input features for model training. And we divide the training and testing datasets in two different ways. In these two ways, the testing datasets are the last 20% of datasets in each condition, and the datasets in the last condition respectively. The evaluation of these models’ performance is exerted with Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R2). The MAE of these three models are SVR: 0.8096, ANN: 0.4984, GRU: 0.7289 in the trained working conditions, and SVR: 1.0946, ANN: 0.7878, GRU: 0.7837 in the untrained working conditions. The results of MAE show that the performance of the ANN is the best among the three models when tested with the trained and untrained test datasets, and the same conclusion can be got from the R2 and RMSE. Moreover, the accuracies of these models are lower when the tested dataset is collected in new working conditions. According to the results above, ANN may be the preferred method for vehicle drivers’ facial air temperature prediction.

INDEX TERMS Artificial intelligence, control nonlinearities, heating, HVAC control, temperature control, temperature measurement, thermal engineering, thermal comfort evaluation.

I. INTRODUCTION

With the development of air-conditioning technology, automatic air conditioning based on thermal comfort is getting more and more attention. The original control method of adjusting and controlling the indoor temperature manually is no longer suitable for future needs. Accurate thermal comfort prediction is essential for controlling an automatic air conditioning system and improving drivers’ thermal comfort [1], [2]. Various methods were developed to model the thermal comfort of car occupants [3]–[10]. The most frequently considered method is the Predicted Mean Vote (PMV) proposed by Fanger and O. P. [5]. K. Matsunaga et al. used the Average Equivalent Temperature (AET) to calculate the PMV and assess the overall average thermal comfort [6]. The calculation of the AET model is mainly based on the surface area of three human body regions: head (0.1), abdomen (0.7), feet (0.2). However, this model failed to illustrate the different local thermal sensitivity of each body part. Y. Taniguchi et al. developed an evaluation method, partial Thermal Sensation Votes (TSV), based on subjective thermal comfort research which only took facial thermal sensation into consideration because of its significant impact on the overall thermal comfort [4]. Hagino and Junichiro improved the partial TSV to whole body TSV by giving different weight values to each part of the human body: forehead (0.42), upper arm on window side (0.38), thigh on window side (0.20), instep on window side (0.28) [11]. It can be concluded from the coefficients of partial TSV and whole body TSV that the face is most sensitive to thermal conditions and contributes the most to thermal comfort. This means that an accurate head temperature prediction model is very important for improving the comfort of the vehicle driver [12]. Machine Learning (ML)
algorithms are appropriate data-driven method of obtaining head temperature [3], [13]–[20]. Many former researchers have applied these methods to predict the temperature in other fields [21]–[23]. Xiao et al. used the machine learning methods to predict the Sea surface temperature with a high accuracy [24]. Shwetha et al. employed the ANN approach to predict the land surface temperature under cloudy conditions precisely [25]. Czerwinski et al. successfully estimated the temperature of Brushless Direct Current (BLDC) [26]. Therefore, it is possible that the machine learning methods are used to predicted the head temperature based on the data of other dimensions. This application of ML methods is also called “data-driven soft sensors” [26], [27], which can give the head temperature as a non-intrusive method [28], [29]. Due to the non-linearity that the machine learning method has, it is very advantageous to deal with such non-linear system with large time lag as the air conditioning system. The machine learning model establishes a mapping between input data and output data. To sum up, with these methods mentioned above, the head temperature can be sensed and used for air conditioning control. Although ML has been universally used in many fields [30]–[32], these studies didn’t predict the head temperature of vehicle drivers which is of particular importance for thermal comfort control.

The automatic control of air conditioning system is usually based on the indoor temperature sensor which is generally set beneath the dashboard. Therefore, there is no way for it to reflect the thermal comfort of human body precisely. Researchers often empirically add or minus a constant value to the value of the indoor temperature sensor to approach the head temperature. Although this method is simple to operate, it is not very convincing and precise enough. For a car with a small compartment, the accuracy may be acceptable, but for a slightly larger car, the temperature difference between the head temperature and the indoor temperature sensor is more complicated. A model with higher accuracy and being able to map this complex relationship is needed. Plenty of researches suggest that data-driven models and physics-based models can meet the need [11], [13], [16], [33], [34]. Although physics-based models are helpful for understanding its internal mechanism, these models usually need a lot of computing power and could not be used for real-time sensing and controlling [35]. Whereas the data-driven models such as SVR, ANN, and GRU will build a mapping between the required input and the head temperature without knowing the physical processes of the mechanic systems [31], [36].

The objective of this study is not only to choose the best data-driven model and explore how much do the Support vector regression (SVR), Artificial Neural Network (ANN), Gated Recurrent Unit (GRU) models overperform the empirical methods, but also to figure out how do these models perform in trained and untrained working conditions. In this paper an experiment is conducted to collect the data of head temperature and other related dimensions. 12 datasets of heating working conditions under different ambient temperatures and heating power are chosen for the head temperature prediction. We employ the SVR, ANN, and GRU models [25], [37]–[39] and optimize them by carefully selecting their hyperparameters. These models are trained with datasets of all working conditions, whereas the test datasets are split to trained and untrained working conditions. The testing datasets of trained working conditions are collected in the same working conditions with the training datasets. Their best performances under trained and untrained working conditions are also evaluated and compared to choose the most appropriate one.

The rest parts of this paper are organized in the structure as follows. In Section 2, the process of data acquisition and preprocessing is given. Whereas Section 3 describes the structure of SVR, ANN, GRU models and their optimization. Section 4 demonstrates the results of these models and discussions. In the end, Section 5 is the conclusion of the whole paper.

II. METHODS

Figure 1 presents the process flow of optimizing and evaluating the SVR, ANN, GRU models. The details of each method and evaluation indicator are illustrated in the following subsections.

A. SUPPORT VECTOR REGRESSION (SVR)

Support vector machine is a popular machine learning technology whose calculation depends on the kernel function [39]. SVM itself is applied to two classification problems. In this article, an important variation of SVM in regression, namely nonlinear SVR, is used. The basis of the theory is to use the kernel function to transform the data to higher dimensions, and then find the best function (f(X) or hyperplane) so that the deviation of the function from the target does not exceed the value of $\varepsilon$ ($\varepsilon$), and make sure the function as flat as possible. The (1) is used to predict the target value, and the coefficients are obtained by minimizing the (2) Lagrangian equation while satisfying the constraint conditions in (3) [39].

$$f(x) = \sum_{i=1}^{n} (a_i^0 - a_i) k(x_i, x) + b \quad (1)$$

$$L(a) = 0.5 \sum_{i=1}^{n} \sum_{j=1}^{n} (a_i - a_j^0) (a_j - a_j^0) k(x_i, x_j) + \varepsilon \sum_{i=1}^{n} (a_i + a_i^0) - \sum_{i=1}^{n} y_i (a_i - a_i^0) \quad (2)$$

$$\sum_{i=1}^{n} (a_i - a_i^0) = 0$$

$$0 \leq a_i \leq C$$

$$0 \leq a_i^0 \leq C \quad (3)$$

In these equations, $x_i$ is the support vector from the input value, $n$ is the number of support vectors, $k$ is the kernel function, which can be Gaussian, linear or polynomial, and $y_i$ is the target value.

B. ARTIFICIAL NEURAL NETWORK (ANN)

Among all the types of ANN, the Backpropagation Neural Network (BPNN) was chosen to predict the head temperature
because the BPNN can adapt to and express the nonlinear dynamic feature effectively, and achieve efficient learning and accurate prediction with Multilayer Perceptron (MLP). This neural network is the simplest of the neural network algorithms and consists of an input layer, several hidden layers, and an output layer. The network structure is shown in Fig. 2. In the input layer, in this study, there are 6 neurons corresponding to the 6 main input vectors which will be given in detail in the section of Data preprocessing. These 6 vectors have the most significant impact on the head temperature and are mutually independent, so they can be used for prediction. The numbers of hidden layer and hidden layer neurons were carefully selected and optimized in this ANN network. There is only one output layer in this study, and its output value is the
predicted head temperature value. The calculation process of each node in the ANN neural network is shown in the Fig. 3.

**FIGURE 3.** The calculation process of ANN.

The **h**1 refers to the first neuron of the hidden layer, in which **f** is the transfer function. While **l** refers to the output layer, in which **f**' is the transfer function of the output layer. **X** is the input, and **W** and **b** are the connection weight and deviation weight of the hidden layer and output layer. The output calculation process of other neurons in the hidden layer is the same as that of the first neuron. The final output **l** can be got after a series of calculations in the input layer, the hidden layer, and the output layer.

According to the error between the target output and the actual output, the neural network will use a specific update algorithm to update its own weight. When the loss reaches the maximum tolerance range or the number of iterations reaches the maximum, the neural network training will be completed.

**C. RECURRENT NEURAL NETWORK (RNN) AND GATED RECURRENT UNIT (GRU)**

Before introducing the GRU network, we need to have a general picture of the RNN model, because the GRU network is an updated version of the RNN and they share the same basic framework. The structure of traditional RNN is shown in Fig. 4. The update equation of the hidden state **h** is as follows:

\[
\mathbf{h}_t = \tanh(\mathbf{U} \mathbf{x}_t + \mathbf{W} \mathbf{h}_{t-1} + \mathbf{b})
\]  

(4)

Tanh is the transfer function named the hyperbolic tangent function, while **b** is the deviation vector. The **U** is the weight matrix of the cell state and the **W** is that of the input. In the traditional RNN models, **X** is the input vector at time **t**, **h** is the final output of the RNN cell at time **t**.

**FIGURE 4.** The traditional RNN structure.

In the last step, the gate value of the reset gate was multiplied by the new **c**' and the **Z** was multiplied by **c** - 1, and then the results of the two were added together to obtain the final hidden state output **c**. In the GRU model, if the gate value tends to 1, the output will be the new **c**; and if the gate value tends to 0, the output will be the **c** - 1 of the previous time step. The equations are given by the authors in [36] and they are replicated in (5) - (9):

\[
r_t = \sigma(W_{x_1} \mathbf{x}_t + W_{h_1} \mathbf{h}_{t-1} + \mathbf{b}_r)
\]

(5)

\[
z_t = \sigma(W_{x_2} \mathbf{x}_t + W_{h_2} \mathbf{h}_{t-1} + \mathbf{b}_z)
\]

(6)

\[
c'_t = \tanh(W_{x_3} \mathbf{x}_t + W_{h_3} (r_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_c)
\]

(7)

The framework of the GRU model is shown in the Fig. 5.
The air outlet has 9 gears which control the wind speed of the outlet from the outlets near the Feet. The wind speed of the face, whereas in the Feet blowing mode, the hot air goes opposite. In the blowing mode to only Feet blowing and Face blowing respectively according to driving habits because the hot air goes upward naturally and the cold air goes opposite. In the heating mode, we set the ambient temperature was set to 5 °C, 0 °C, −5 ° and −10 °C respectively, whereas, in the cooling mode, it was set to 25 °C, 30 °C, 35 °C, and 40 °C. Additionally, the intensity of solar radiation is also simulated by controlling the Xenon lamp. The simulated solar radiation intensity ranges from 0 W/m² to 400 W/m². Above all, these working conditions were different in such dimensions as the ambient temperature, the wind speed gear, solar illuminance, and the rotating speed of the compressor, etc.

The experimenters arranged 37 sensors in the car, covering almost all data related to head temperature. The experimental data is collected once every 6 s, and there are a total of 6526 sets of data in all these conditions. These dimensions of data can well reflect the heating and cooling performance of the air conditioning system. However, the parameters of data-driven models in the heating and cooling working conditions are different because of the difference in the wind blowing mode, whereas, their developing methods are the same. Therefore, in this article, we only discuss the model in heating conditions.

D. EVALUATION OF THE MODEL PERFORMANCE

Root Mean Square Error (RMSE) (10), the root mean square error is the square root of the ratio of the sum of the squares of the deviation between the predicted value and the true value to the number of predictions m. It is used to measure the deviation between the observed value and the true value. Mean Absolute Error (MAE) (11) is the average value of absolute error, which can better reflect the actual situation of the predicted value error. Coefficient of determination ($R^2$) (12), which reflects the proportion of all the variation of the dependent variable that can be explained by the independent variable through the regression relationship. The equations of the above three determination coefficients are shown below. They are the key criteria for model evaluation. The performances of models are all evaluated with these three indicators in this study.

\[
RMSE(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2} \quad (10)
\]

\[
MAE(X, h) = \frac{1}{m} \sum_{i=1}^{m} |\hat{y}_i - y_i| \quad (11)
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (12)
\]

In these equations, $\hat{y}_i$ is the predicted value and the $y_i$ is the raw data. $\bar{y}$ is the average value of raw data.

E. PROGRAMMING ENVIRONMENT

The modeling and evaluating process of our research were mainly performed in Python 3.6. The data processing is conducted with the SPSS and such packages of the Python software as NumPy [42], Pandas [43], and Scikit-Learn [44]. Our SVR, ANN, and GRU networks were established with Keras based on the Google TensorFlow [15].

IV. RESULTS AND DISCUSSION

A. DATA PREPROCESSING

Through data analysis methods such as Pearson correlation analysis, data of several extremely correlated dimensions (such as the temperature of several facial air outlets) were represented by one dimension, and the data that is not related to the head temperature (such as the compressor rotating speed, etc.) were eliminated. Finally, these data were simplified to 7 dimensions, including ambient temperature, vehicle speed, wind speed gear, solar illuminance, temperature of outlet near drivers’ feet, inner temperature sensor, and head temperature. As is mentioned in Section III, there are a total of 6526 samples collected in all these conditions and there are 6 input data and 1 output data namely the head temperature in each sample. To sum up, the input matrix size is 6526 * 6, and the output matrix size is 6526 * 1.

In order to apply machine learning methods, the data needs to be processed into supervised learning data and divided into training dataset and testing dataset. However, to interpret the performance of these models more precisely, the test data was chosen in two ways. In the first way, as is shown in Fig. 6 (a), we chose the whole dataset of a certain working
eters. The first thing to consider when determining model performance essentially the process of finding the optimal hyperparameters. The development of ANN models is directly related to their hyperparameters. The development of ANN models is the process of finding the optimal hyperparameters. The first thing to consider when determining model performance essentially the process of finding the optimal hyperparameters.

2) DEVELOPMENT OF ANN

All the training and testing data were then normalized into \( [0, 1] \) using (13).

\[
x_{\text{norm}} = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (13)
\]

In this equation, \( x_{\text{norm}} \), \( x_i \), \( x_{\text{min}} \) and \( x_{\text{max}} \) are the normalized, observed, minimum and maximum values of data.

B. MODEL DEVELOPMENT

1) DEVELOPMENT OF SVR

The training effect of the SVR model mainly depends on three parameters, the insensitive loss coefficient \( \epsilon \), the width coefficient \( \gamma \), and the penalty coefficient \( C \). If \( \epsilon \) is too large, the number of support vectors will be small, which may lead to the model being too simple and the learning accuracy is insufficient; if \( \epsilon \) is too small, the regression accuracy will be higher, but the model may be too complicated. The width coefficient \( \gamma \) reflects the degree of correlation between the support vectors. The larger the \( \gamma \), the fewer the support vectors, the looser the connection between the vectors, the simpler the model, and the unsatisfactory training effect. The smaller the \( \gamma \), the more support vectors, and the more closely the relationship, the more complex the model, and the generalization performance may not be ideal. The higher the penalty coefficient \( C \) is, the easier it is to overfit. The smaller \( C \) is, the easier it is to underfit. In this paper, the gradient optimization algorithm of GridsearchCV [45] is used to optimize the SVR model [46]. It can automatically calculate the optimal combination of parameters within a given range and give the optimal results. Finally, the optimal parameters of the model are \( C = 0.285 \), \( \gamma = 0.42 \), epsilon = 0.001.

2) DEVELOPMENT OF ANN

The performance of the ANN models is directly related to their hyperparameters. The development of ANN models is essentially the process of finding the optimal hyperparameters. The first thing to consider when determining model hyperparameters is the transfer function. The three most used transfer functions in neural network models are Sigmoid, Tanh, and Relu. The output of Sigmoid is not zero-mean-value, therefore, the input of the neurons in the subsequent layer is non-zero-mean signals, which will slow down the gradient training. This transfer function is generally used in two classification problems. The output of the Tanh function is 0-mean, so Tanh is better than Sigmoid in practical applications. Relu will make the output of some neurons to be 0, which reduces the complexity of the network and the interdependence of parameters, alleviating the occurrence of over-fitting problems, but it will cause neuron necrosis.

In several ANN models with different structures, the transfer function Relu, Tanh, and Sigmoid were employed and compared respectively. The models were trained ten times on each condition and the prediction results of the model were averaged. Under the same model structure, different transfer functions were evaluated through many comparative experiments. Experimental results show that the best model effect can be achieved when the transfer function is selected as Relu. Therefore, the activation function of Relu was finally used in the model.

Additionally, optimization algorithms and loss functions can also affect the training effect of the model to a large extent. Stochastic Gradient Descent (SGD) was initially considered, but the SGD with a fixed learning rate will cause severe oscillations in the cost function. Although it can reach a minimum value, it takes longer than other algorithms and may be trapped at the saddle point. Consequently, the SGD is not the preferred optimization algorithms in this study. In order to converge faster, an algorithm that has an adaptive learning rate is needed. The latest two adaptive learning rate algorithms are Adaptive momentum estimation (Adam) and Root Mean Square prop (RMSprop). Adam adds bias-correction and momentum to RMSprop. Therefore, theoretically speaking, as the gradient becomes sparser, Adam has a better effect than RMSprop. Overall, Adam is the best choice.

The commonly used loss functions for training neural networks are mainly Mean Absolute Error (MAE) and Mean Square Error (MSE). After cross-applying optimization functions Adam and RMSprop and loss function MAE and MSE to the model, it is found through a lot of comparative experiments that the model combining Adam and MSE performs better, followed by Adam and MAE, RMSprop and MSE, and the combination of RMSprop and MAE performed worst.

Except for improving the training effect, we should also simplify the network structure as much as possible to reduce the cost of online computing power. Therefore, it is necessary to choose between a better training result and a smaller number of neurons. With respect to the structure of the hidden layer of the neural network, we mainly consider the single-layer and double-layer cases. In order to compare their performance, we tested the single hidden layer and double hidden layer neuron models, and selected their structures as the single hidden layer structure of neurons from 5 to 20 and...
double hidden layer structure of neurons from 1-1 to 15-15. The result (as shown in the Fig. 7 (a) and Fig. 7 (b), in which the upper and lower dotted lines represent the best and worst performance of the models respectively) suggests that as the number of neurons increases, the accuracy of the double hidden layer model shows an upward trend. And it has a better prediction effect than a single hidden layer model whose MAE is 0.2 °C higher than the double hidden layer model on average. In addition, the gap between upper and lower dotted lines in Fig. 7 (b) is relatively narrower, which represents a more stable performance of the double hidden layer model. And the best performance was reached when the network structure is 14-14.

![FIGURE 7. The performance of Single-layer ANN (a) and Double-layer ANN (b).](image)

The training process of this model is equivalent to the offline training process. Therefore, when the time cost is acceptable, the increase in computing caused by the larger number of iterations can be ignored. Consequently, 200 epochs are selected in this project. This number of iterations ensures that all models converge.

3) DEVELOPMENT OF GRU

As for the hyperparameter optimization in GRU, the trial-and-error method was employed. Through a lot of experiments, we concluded that the single-layer GRU network is better than the multi-layer GRU in predicting the head temperature. Like the ANN model, the GRU model also needs to choose an optimizer, loss function, number of neurons, etc. When the number of neurons in the GRU network is 2, the two loss functions MSE and MAE and the two optimizers Adam and RMSprop are cross-experimented. The experiment was run 10 times in each case and the evaluation parameters were averaged to eliminate errors. Experimental results shown in Table 1 is tested by the trained working condition, which proves that in the GRU neural network, the best combination is still Adam and MSE.

![FIGURE 8. Performance of GRU with different network structure.](image)

| Optimizer | MSE | MAE | MSE | MAE |
|-----------|-----|-----|-----|-----|
| Adam      | 0.7464 | 0.7738 | 0.8660 | 0.9006 |
| RMSprop   | 0.9603 | 0.9559 | 0.9497 | 0.9483 |

Training was performed when the number of neurons in a single layer of the GRU neural network were 1, 2, 4, 8, and 16 respectively. The training results are shown in Fig. 8, in which the upper and lower dotted lines represent the best and worst performance of the models respectively. It is found that when there are two neural nodes in the GRU models, it performs the best. And a GRU model with more than two neural nodes perform worse with the rise of neural nodes number.

For ANN and GRU models, another issue that needs to be considered is overfitting. When the loss of test data is much higher than the training loss, chances are that the model has been overfitted. The reason for overfitting is the high model complexity. There are many strategies to avoid overfitting by reducing the complexity of the model. For example, L1 or L2 regularization uses the Lagrangian method to optimize the penalty function formed by the L1 and L2 norms of the weight matrix together with the cost function that needs to be optimized, to control the number of non-zero elements in the model. Additionally adding a dropout of a certain size to randomly deactivate the nodes in the model at a certain ratio is also a useful way to reduce the complexity of the model. In this paper, the overfitting is controlled by adding a dropout value of 0.1.

C. PERFORMANCE COMPARISON AMONG OPTIMIZED SVR, ANN, GRU MODELS, AND THE INDOOR TEMPERATURE SENSOR

After the hyperparameters of several models have been determined, the models were compared in their best performance. According to the experimental results, the best performance of several models can be sorted from good to
TABLE 2. The Best performance of SVR, ANN, GRU in the trained and new condition and the performance of empirical model.

|                  | RMSE     | R2       | MAE     |
|------------------|----------|----------|---------|
| **Trained**      |          |          |         |
| SVR              | 0.9389   | 0.9591   | 0.8096  |
| ANN              | 0.6388   | 0.9811   | 0.4984  |
| GRU              | 0.9074   | 0.9709   | 0.7289  |
| **New**          |          |          |         |
| SVR              | 1.3335   | 0.9176   | 1.0946  |
| ANN              | 0.9519   | 0.9073   | 0.7878  |
| GRU              | 0.9832   | 0.9011   | 0.7837  |
| **Empirical model** | 3.2992 | 0.9064   | 2.9774  |

TABLE 3. The hyper-parameters of optimized SVR, ANN, GRU.

|         | Neuron numbers | Transfer function | Optimization algorithms | Loss function | C    | \( \gamma \) | \( \varepsilon \) |
|---------|----------------|-------------------|-------------------------|---------------|------|-------------|-----------------|
| SVR     | --             | --                | --                      | --            | 0.285| 0.42        | 0.001           |
| ANN     | 5-5            | Tanh              | Adam                    | MSE           | --   | --          | --              |
| GRU     | 2              | Tanh              | Adam                    | MSE           | --   | --          | --              |

Comparing the MAE curves of the ANN model and the GRU model, it was found that the hidden layer structure of the ANN model between 5-5 and 15-15 has basically stabilized at the best performance. The performance evaluation parameters of the ANN model with a structure of 14-14 in these two working conditions are respectively: RMSE: 0.9519, R2: 0.9073, and MAE: 0.7878 in the new working condition; RMSE: 0.6388, R2: 0.9811, and MAE: 0.4984 in the trained working condition. When the number of neurons in the GRU model is 2, its model performance is basically the best. On this condition, the model performance parameters are RMSE: 0.9074, R2: 0.9709, and MAE: 0.7289 on the trained condition and RMSE: 0.9832, R2: 0.9011, and MAE: 0.7837 on the new condition.

The evaluation parameters of the SVR model are RMSE = 0.9389, R2 = 0.9591, MAE = 0.8096 on the trained working condition; RMSE = 1.3335, R2 = 0.9176, MAE = 1.0946 on the new working condition.

Comparing the best performance of the three models, it is found that the best performance of the ANN model is better than the GRU model and the SVR model in this subject. After comprehensively considering the parameter requirements and performance requirements of the project, it was decided to adopt the ANN model as the final model. Considering the minimum parameter amount for the best performance, the structure of the ANN model in this study is selected as a 5-5 double hidden layer structure. As a variant of LSTM [47], GRU’s advantages are mainly reflected in memorizing and analyzing the characteristics of long-term data. Since the predicted output of the current time step in this subject is estimated based on input data of other dimensions at the previous and current time steps, the advantages of the GRU model over ANN in long-term prediction are not reflected. Therefore, this result is understandable.

In order to discuss the effectiveness of these models, we also need to compare their performance differences with the previous empirical methods. Before we introduced the facial air temperature prediction algorithms into the control of automatic air conditioning system, the control of it was based only on the indoor temperature sensors which are usually placed far from drivers’ head. Therefore, the difference between the indoor temperature sensor data and the head temperature data is also evaluated using the performance evaluation coefficient. The result is RMSE: 3.2992, R2: 0.9064, MAE: 2.9774, which is much worse than the predicted data. The performance comparison between SVR, ANN, GRU in the trained and new condition and the empirical model is shown in Table 2.

D. IMPACTS OF WORKING CONDITION

As we can see in Table 2, the performance of these three models in the new working condition is relatively poor...
compared with the trained working condition. However, in the process of product design of a real vehicle prediction model, all the data of common working conditions will be collected for model training. In that case, the ANN models will be able to master most of the working conditions.

**E. PREDICTION DATA OF THE FACIAL AIR TEMPERATURE**

Fig. 10 shows the raw data of head temperature and the predicted temperatures of the three data-driven models for the 2 testing datasets in around 1300 time-step. The results demonstrate ANN model can precisely predict the head temperature and reflect the trend of head temperature changing over time.

**V. CONCLUSION**

To develop a head temperature prediction model with high accuracy and low compute power cost and apply it in the control strategy of vehicle drivers’ thermal comfort, we selected and optimized the traditional machine learning algorithm SVR, the traditional ANN model, and the variant of RNN network, GRU. The performances of the models were compared and evaluated. The best performance of the ANN model is RMSE: 0.6388, R2: 0.9811, MAE: 0.4984, and the best performance of GRU is RMSE: 0.9074, R2: 0.9709, and MAE: 0.7289 on the trained condition. The performance of SVR is RMSE: 0.9389, R2: 0.9591, MAE: 0.8096. Based on the comparison of these models’ performance and the limitation of parameters’ number, the ANN model with a 5-5 structure was selected for head temperature prediction in this study. For it performs close to the best ANN network structure (14-14) and has a significantly smaller number of parameters than the best structure as shown in the Fig. 7(a). The hyper-parameters of the proposed models are shown in the Table 3. The next step of our study is to complete the predictive control strategy of the driver’s head temperature in the passenger compartment of the car and collect more data on other working conditions for further model training. The following conclusions are summarized from this study:

1. High-precision temperature prediction can be achieved by inputting data in a limited number of dimensions in the passenger compartment.
2. The prediction performance of ANN is better than that of GRU model and SVR model when tested with the trained working condition.
3. These three models perform better when tested with the trained working condition than the new working condition.

In this study, a variety of machine learning methods were employed to predict the head temperature, which provides a computing-power saving online predicting method for air conditioning system control based on thermal comfort. With a slight modification, this model can also be used to predict values other than head temperature such as the abdomen temperature, which provides the possibility to comprehensively predict and control other indicators that affect drivers’ thermal comfort. In our future work, we will conduct an end-end test to examine our models and combine them with the control strategies in a real electric vehicle. Additionally, we will collect the datasets of more working scenarios for further training.
REFERENCES

[1] J. Kim, S. Schiavon, and G. Brager, “Personal comfort models—A new paradigm in thermal comfort for occupant-centric environmental control,” Building Environ., vol. 132, pp. 114–124, Mar. 2018.

[2] W. Li, J. Zhang, and T. Zhao, “Indoor thermal environment optimal control for thermal comfort and energy saving based on online monitoring of thermal sensations,” Energies, vol. 197, pp. 57–67, Aug. 2019.

[3] Y. D. Guan, M. H. Hosni, B. W. Jones, and T. P. Gielda, “Literature review of the advances in thermal comfort modeling,” ASHRAE Trans., vol. 109, no. 2, pp. 908–916, 2003.

[4] Y. Taniguchi, H. Aoki, K. Fujikake, H. Tanaka, and M. Kitada, “Study on car air conditioning system controlled by car occupants’ skin temperatures—Part I: Research on a method of quantitative evaluation of car occupants’ thermal sensations by skin temperatures,” SAE Tech. Paper, USA, Tech. Rep. 920170, 1992.

[5] O. P. Fanger, Thermal Comfort: Analysis and Applications in Environmental Engineering, 1970. Copenhagen, Denmark: Danish Technical Press, 1972, p. 244.

[6] K. Matsunaga, F. Sudo, S. I. Tanabe, and T. L. Madsen, “Evaluation and measurement of thermal comfort in the vehicles with a new thermal manikin,” in SAE, 7th Int. Pacific Conf., 1993, p. 931958.

[7] E. Arens, H. Zhang, and C. Huizenga, “Partial- and whole-body thermal sensation and comfort—Part I: Uniform environmental conditions,” J. Thermal Biol., vol. 31, nos. 1–2, pp. 53–59, Jan. 2006.

[8] E. Arens, H. Zhang, and C. Huizenga, “Partial- and whole-body thermal sensation and comfort—Part II: Non-uniform environmental conditions,” J. Thermal Biol., vol. 31, nos. 1–2, pp. 60–66, Jan. 2006.

[9] K. Irshad, A. I. Khan, S. A. Irfan, M. M. Alam, A. Almalawi, and M. H. Zahir, “Utilizing artificial neural network for prediction of occupants thermal comfort: A case study of a test room fitted with a thermo-electric air-conditioning system,” IEEE Access, vol. 8, pp. 99709–99728, 2020.

[10] H. Zhang, C. Huizenga, E. Arens, and D. Wang, “Thermal sensation and comfort in transient non-uniform thermal environments,” Eur. J. Physiol., vol. 92, no. 6, pp. 33–72, 2004.

[11] Z. Wang, H. Onodera, and R. Matsuhashi, “Proposal of relative thermal sensation: Another dimension of thermal comfort and its investigation,” IEEE Access, vol. 9, pp. 36266–36281, 2021.

[12] M. Jia, J.-H. Choi, H. Liu, and G. Susman, “Development of facial-skin temperature driven thermal comfort and sensation modeling for a futuristic application,” Building Environ., vol. 207, Jan. 2022, Art. no. 108479.

[13] D. Li, C. C. Menassa, and V. R. Kamat, “Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography,” Energy Buildings, vol. 176, pp. 246–261, Oct. 2018.

[14] S. Shamshirband, M. Fathi, A. Dehzangi, A. T. Chronopoulos, and H. Almeijead-Rokny, “A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues,” J. Biomed. Informat., vol. 93, Jan. 2020, Art. no. 101272.

[15] M. T. Sattari, H. Apaydin, and S. Shamshirband, “Performance evaluation of deep learning-based gated recurrent units (GRUs) and tree-based models for estimating ETo by using limited meteorological variables,” Mathemat. vol. 8, no. 6, p. 972, Jun. 2020.

[16] G. Zhang, S. S. Band, C. Jun, S. M. Bateni, H.-M. Chuang, H. Turabieh, M. Mafarja, A. Mosavi, and M. Moslehpour, “Solar radiation estimation in different climates with meteorological variables using Bayesian model averaging and new soft computing models,” Energy Rep., vol. 7, pp. 8973–8996, Nov. 2021.

[17] Lee, Choi, Choi, and Kim, “Development of a data-driven predictive model of clothing thermal insulation estimation by using advanced computational approaches,” Sustainability, vol. 11, no. 20, p. 5702, Oct. 2019.

[18] J. W. Moon, “Performance of ANN-based predictive and adaptive thermal-control methods for disturbances in and around residential buildings,” Building Environ., vol. 48, pp. 15–26, Feb. 2012.

[19] A. Martynenko and N. N. Misra, “Machine learning in drying,” Drying Technol., vol. 38, nos. 5–6, pp. 596–609, Apr. 2020.

[20] C.-C. Young and W.-C. Liu, “Prediction and modelling of rainfall–runoff during typhoon events using a physically-based and artificial neural network hybrid model,” Hydrolog. Sci. J., vol. 60, no. 12, pp. 2102–2116, Dec. 2015.

[21] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” Comput. Sci., pp. 1–15, Jun. 2014. Online: Available: https://arxiv.org/abs/1406.1078

[22] B. Schiklof, A. J. Smola, R. C. Williamson, and P. L. Bartlett, “New support vector algorithms,” Neural Comput., vol. 12, no. 5, pp. 1207–1245, May 2000.

[23] B. Schiklof and A. J. Smola, Learning With Kernels—Support Vector Machines, Regularization, Optimization and Beyond. Cambridge, MA, USA: MIT Press, 2001.

[24] J. Zhang, Y. Zhu, X. Zhang, M. Ye, and J. Yang, “Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas,” J. Hydrol., vol. 561, pp. 918–929, Jun. 2018.

[25] S. van der Walt, S. C. Colbert, and G. Varoquaux, “The NumPy array: A fundamental representation model,” IEEE Access, vol. 6, no. 1, pp. 2102–2116, Dec. 2015.

[26] X. Zhang et al., “Prediction of Vehicle Driver’s Facial Air Temperature With SVR, ANN, and GRU,” IEEE Access, vol. 8, pp. 131534–131544, 2020.

[27] C. Xiao, N. Chen, C. Hu, K. Wang, Z. Xu, Y. Cai, L. Xu, Z. Chen, and J. Gong, “A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data,” Environ. Model. Softw., vol. 120, Oct. 2019, Art. no. 104557.

[28] H. R. Shwetha and D. N. Kumar, “Prediction of land surface temperature under cloudy conditions using microwave remote sensing and ANN,” in Proc. Int. Conf. Water Resour., Coastal Ocean Eng. (ICWRCOE), Karnataka, India, vol. 4, Jan. 2015, pp. 1381–1388.

[29] D. Czerwinski, J. Géca, and K. Kolano, “Machine learning for sensorless temperature estimation of a BLDC motor,” Sensors, vol. 21, no. 14, pp. 4655, Jul. 2021.

[30] P. Kadlec, B. Gabrys, and S. Strødt, “Data-driven soft sensors in the process industry,” Comput. Chem. Eng., vol. 33, no. 4, pp. 795–814, Apr. 2009.

[31] M. Jia, J.-H. Choi, H. Liu, and G. Susman, “Development of facial-skin temperature driven thermal comfort and sensation modeling for a futuristic application,” Building Environ., vol. 207, Jan. 2022, Art. no. 108479.

[32] L. Xian, T. Guan, K. Fan, and Q. Sun, “Novel double layer BiLSTM minor soft fault detection for sensors in air-conditioning system with KPCA reducing dimensions,” J. Building Eng., vol. 44, Dec. 2021, Art. no. 102950.

[33] Y. Jin, Y. Ying, J. Li, and H. Zhou, “Gas path fault diagnosis of gas turbine engine based on knowledge-driven artificial intelligence algorithm,” IEEE Access, vol. 9, pp. 108932–108941, 2021.

[34] R. Kannan and M. S. Roy, “AC cooling time prediction using common representation model,” IEEE Access, vol. 8, pp. 131534–131544, 2020.
[46] P. S. Yu, S. T. Chen, and I. F. Chang, “Support vector regression for real-time flood stage forecasting,” J. Hydrol., vol. 328, nos. 3–4, pp. 704–716, Sep. 2006.

[47] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

**XIAOHAN ZHANG** received the bachelor’s degree from Beijing Jiaotong University. He is currently pursuing the master’s degree with the Beijing Institute of Technology. His research interests include air conditioning system design, especially in control strategy, system engineering, and temperature prediction.

**YICHUN WANG** was born in Jilin, in 1963. He received the bachelor’s degree from the Development Department, Daqing Petroleum Institute, in 1986, the master’s degree from the School of Machinery and Vehicles, Beijing Institute of Technology, in 1993, and the Ph.D. degree from the Power Machinery Manufacturing Department, Moscow State Bauman University of Technology, in 2000.

He is the Special Commissioner for Science and Technology, Beijing Institute of Technology, Guangdong, and the Deputy Chief Designer of a Foreign Trade Special Vehicle at China Ordnance. He holds more than 20 national patents. He has published more than 80 papers, including more than 40 papers retrieved by SCI and EI. His research interests include new energy power system technology, new energy vehicle heat pump air conditioning technology, power battery, and power system extreme environment adaptation technology.

Dr. Wang is a Senior Member of the Chinese Ordnance Engineering Society and a member of the Chinese Society of Automotive Engineering. In 2007, he won the Beijing Institute of Technology Special Award for Technological Invention, ranking 4th, the Second Prize of Technological Inventions of the Commission of Science, Technology, and Industry for National Defense, ranking 4th, and the First Prize of China North Industries Group Corporation for Science and Technology Progress, ranking 8th. In 2009, he won the Second Prize of the State Council for Technological Inventions, ranking 4th. In 2013, he won the Second Prize of the Science and Technology Progress of the Ministry of Industry and Information Technology, ranking 7th. He is a Power System Review Expert at the Chinese Academy of Ordnance Sciences.

**HONGZENG JI** received the bachelor’s degree from the China University of Mining and Technology and the master’s degree from the Beijing Institute of Technology, where he is currently pursuing the Ph.D. degree. His research interests include air conditioning system design, air-conditioning system control strategy, and novel CO2 air-conditioning system design.

**YAWEN LI** received the bachelor’s and master’s degrees from the Beijing Institute of Technology, where he is currently pursuing the Ph.D. degree. His research interests include air-conditioning system design, control strategy, system engineering, and temperature prediction.

**XIUGHUI DUAN** received the bachelor’s degree from the Taiyuan University of Technology. He is currently pursuing the Ph.D. degree with the Beijing Institute of Technology. His research interest includes air conditioning system design, especially in control strategy, system engineering, and air-conditioning compressor design.

**FEN GUO** works at the State Key Laboratory of Electric Vehicles, Beijing Institute of Technology. His research interests include intelligent vehicles, neural networks, and fuzzy control.

---

**IEEE Access**

**X. Zhang et al.: Prediction of Vehicle Driver’s Facial Air Temperature With SVR, ANN, and GRU**

**20222**

**VOLUME 10, 2022**