Sensor applications

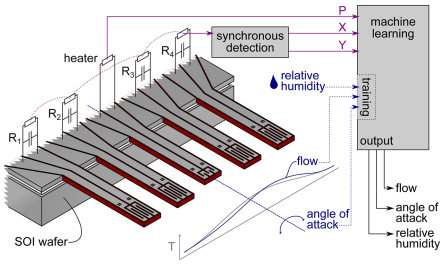
A Machine Learning Enhanced MEMS Thermal Anemometer for Detection of Flow, Angle of Attack, and Relative Humidity

T.L. Hackett, J. Choi, R. G. P. Sanders, T. E. vd. Berg, D. Alveringh*, and J. Schmitz**

MESA+ Institute for Nanotechnology, University of Twente, 7522 NB Enschede, Netherlands
**Member, IEEE*
***Senior Member, IEEE*

Manuscript received 7 June 2024; accepted 20 June 2024. Date of publication 24 June 2024; date of current version 5 July 2024.

Abstract—By optimizing machine learning (ML), the accuracy of a thermal anemometer has been improved (511%) when compared to conventional linear regression. In addition, ML has extended the functionality allowing for additional angle of attack and humidity information to be determined. The miniature sensor (0.16 cm^2) has been fabricated with a straightforward silicon on insulator (SOI) fabrication procedure. The sensor paired with ML could offer a cost-effective, small, and reliable solution for monitoring air in industrial and agricultural sensor grid applications, such as data centers and greenhouses. This proof of principle shows that thermal anemometers can have their accuracy and functionality enhanced through ML, enabling the estimation of multiple physical parameters with a single sensor.



Index Terms—Sensor applications, angle measurement, flow sensor, humidity, machine learning (ML), thermal anemometer.

I. INTRODUCTION

Consistent ventilation and air flow in a green house or vertical farm prevents diseases and crop failure [1]. Being able to sense the air speed in a range of 0–1 m s^{-1} over the entire facility could have large benefits in terms of crop yield. Thermal anemometers are a potential candidate for this purpose, as well as other industrial sensor grid applications, such as large data centers.

Thermal anemometers measure the speed of air (v) through heat transfer from a heating element into the flowing medium. Hot-wire anemometers (HWA) measure the power (P) supplied to the heater to maintain the temperature (T) or current (I) of the heater, while calorimetric sensors measure the differential temperature upstream and downstream from the heater. The measurement for both types of thermal sensor also depends on additional parameters, such as the temperature, pressure, and thermal properties of the gas.

Extending the range of anemometers and minimizing sensitivity toward unwanted physical dimensions has been achieved [2]. This is at the cost of bulky and complex sensor designs, by adding additional sensing elements to compensate for other physical parameters through correlation. Most research aims to improve the measurement resolution of air speed sensors, some even measuring air speed in the range of 0–5 m s^{-1} to a precision of 1 cm s^{-1} [3], [4]. Other research advances the topic of anemometry by using advanced fabrication techniques or sensor geometries [5], [6]. This letter aims to explore the possibility of using machine learning (ML) to improve a simple and straightforward sensor for measuring air speeds in the range of 0–2 m s^{-1}, while simultaneously increasing its functionality.

ML has already shown to more accurately calculate flow speeds (v) or gas compositions when compared to conventional calibration and linear fits [7], [8]. However, a gap in current research is the use of ML to estimate multiple physical parameters using different simultaneous measurements from a single device. Previous work has focused on using ML to improve flow measurements [7] or has made anemometers angle-independent through physical means [9], [10]. By applying ML to individually sampled signals from each sensing element, the angle, air speed, and relative humidity (%RH) can be independently extracted without the need for additional design requirements. Furthermore, nonidealities and nonlinearities of thermal flow sensor outputs or inherent thermal hysteretic effects can be tackled using ML.

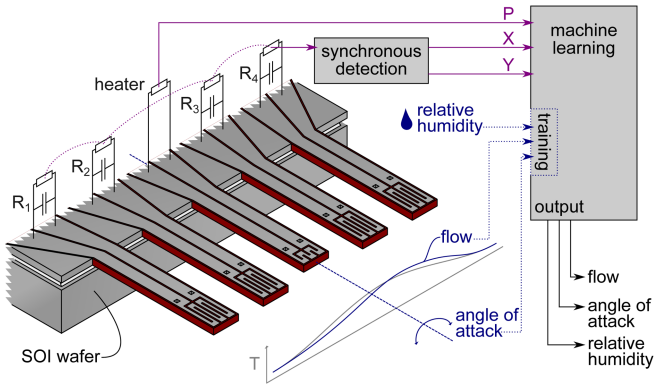


Fig. 1. Schematic of the microfabricated flow sensor with the working principle of ML integration shown. Not shown is a protective bracket around the sensors made from the SOI handle layer.

II. SENSOR DESIGN

The sensor (see Fig. 1) consists of five p-doped silicon resistors: one heating resistor and four measuring resistors placed up- and downstream [11], which can be used for differential Wheatstone bridge readouts.

Corresponding author: T.L. Hackett (e-mail: t.hackett@utwente.nl).
Associate Editor: Chang-Hee Won.
Digital Object Identifier 10.1109/LSENS.2024.3418193

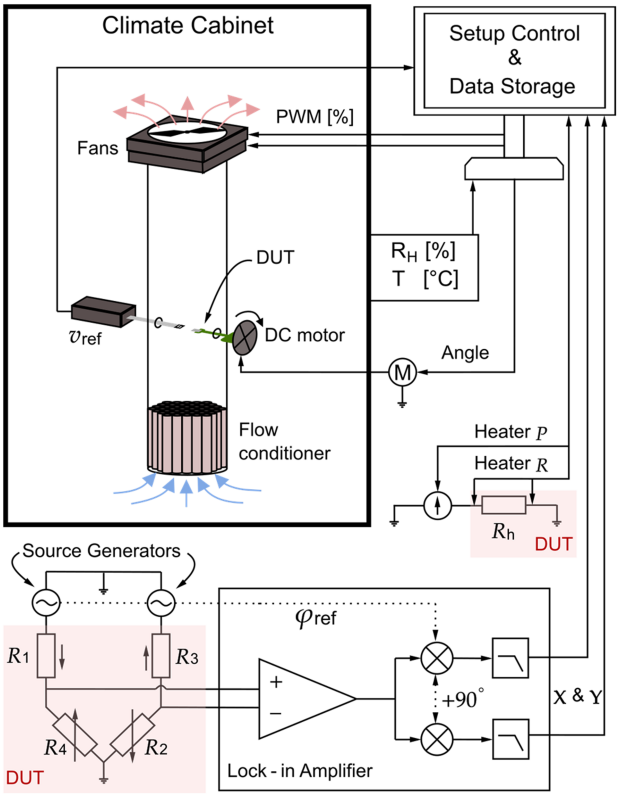


Fig. 2. Schematic of the measurement setup. The measuring resistors and heating element are highlighted and labeled as "device under test." The setup records and stores the sensor dataset, which is later used for ML training and testing purposes.

When flow speed v is applied to the sensor, the two upstream resistors are cooled, while the two downstream resistors are heated, enabling calorimetric flow sensing. The sensor is angle-dependent due to the relative change of position of the sensing elements within the flow induced temperature distribution around the heater.

The heater temperature is kept constant using a proportional-integral controller. Therefore, the heater can additionally be used as a HWA if the supplied electrical power is recorded. As the HWA is approximately cylindrically symmetric with regards to the Angle of Attack (AoA), the HWA flow measurement should be angle-independent. This means that this device combines an angle-dependent calorimetric anemometer and an angle-independent HWA.

As the sensing and heating resistors interface directly with air, they experience a change in impedance due to a change in the relative permittivity of the nearby air. This causes the sensor to be sensitive to %RH changes. The additional output data combined with the calorimetric and HWA output can then be used to train the ML models to find flow speed, AoA, and %RH.

III. EXPERIMENTAL SETUP

The measuring resistors are placed in a Wheatstone bridge configuration and measured using a lock-in amplifier (see Fig. 2). Two synchronized arbitrary waveform generators apply a sinusoidal signal with an amplitude of 250 mV and a frequency of 15.5 kHz over the Wheatstone bridge. A reference signal is provided to the lock-in amplifier which measures an in-phase (X) and quadrature (Y) value.

The dedicated measurement setup is able to apply specific air flows using two pulsewidth modulation controlled fans [12]. The AoA is varied using an automatic rotating stage, while humidity and ambient temperature are maintained by a controlled environment plant growth cabinet (PRC 1200 WL, Hettich Benelux). Applied experimental conditions are recorded in a dataset with 100 datapoints per condition. All of these conditions are cycled and applied to the sensor giving a dataset totaling approximately 300 000 datapoints.

Table 1. Optimized Parameters for K-NN, RF and GB Regressors

Model	Parameter	Flow	Angle	Humidity
K-NN	`n_neighbors`	7	7	9
RF	`n_estimators`	300	400	400
GB	`n_estimators`	400	400	400

Table 2. A Spearman's Correlation Between Physical Input Conditions and Electrical Outputs P, X and Y

Outputs	Physical Input Conditions		
	v [m s^{-1}]	Angle [°]	RH [%]
P	**0.921**	−0.016	−0.019
X	**0.930**	**0.150**	0.044
Y	**−0.724**	−0.027	**0.402**

A. Model Optimization and Training

The power supplied to the heater, as well as the in-phase and quadrature components of the lock-in amplifier, are used to train and test four commonly used ML models, i.e., linear regression (LR), K-nearest neighbors (K-NN) regressor [13], random forest (RF) regressor [14], and gradient boosted (GB) regressor [15], for the estimation of flow speed (v), AoA (θ), and %RH. As a benchmark, this is compared to conventional LR based on just one sensor output, in this case the X component (LR-X). The models are deployed using scikit-learn version 1.2.2 [16] in Python 3.11.3.

The training was performed on a separate consumer grade PC (Windows) with up-to-date hardware including a 12th Gen Intel Core i7-12650H (2.30 GHz) and 15.6 GB of useable RAM and was done using all available cores with float64 encoding.

LR is a model which finds the best linear fit for a dataset by minimizing the least square residuals and works best for linear data. GB regression generates multiple shallow decision-trees (learners) using feature extraction. These learners then build on top of each other sequentially to predict a parameter using the input dataset. RF is also a decision-tree based model, with the distinction that the individual learners are averaged and are independent from each other. Finally, K-NN regression works by finding the numerically nearest known k input values and then assigns the average of the associated output value as the predicted variable. GB, LR and K-NN regressors have the advantage that they don't rely on a linear relationship to make accurate predictions.

The ML models were trained using preprocessed data, which had erroneous zero measurements, as well as apparatus induced offsets removed. The data was then standardized using StandardScaler from scikit-learn. The large dataset was divided into an 80/20 split for training and test sets. A K-Fold cross validation [17] was performed to select for the best model parameters (see Table 1), after which they were evaluated using the mean absolute error (MAE) as an indication of how well each model performed.

IV. RESULTS AND DISCUSSION

A Spearman's rank correlation (see Table 2) shows how much a given variable monotonically correlates with any other variable in a dataset, with 0 indicating no correlation and −1 and 1 showing the maximum negative or positive correlation, respectively.

Table 3. LR Based on Just the In-Phase Component X (LR-X) Compared to ML Models Using P, X, and Y as Inputs

Model	Total Training Time [s]	Average Inference Time [s]	Mean Absolute Error (MAE)		
			v [m s^{-1}]	Angle [°]	RH [%]
LR-X	4.9×10^{-2}	2.2×10^{-9}	0.226	-	-
LR	1.8×10^{-1}	5.8×10^{-9}	0.170	13.1	9.6
GB	7.3×10^{2}	5.3×10^{-6}	0.110	9.6	7.5
K-NN	4.9×10^{-1}	6.4×10^{-6}	0.047	3.9	3.0
RF	1.4×10^{3}	89.5×10^{-6}	0.044	3.6	2.7

From Table 2, it becomes clear that, as expected, flow speed is most correlated with all measured output variables, showing values close to 1 for both P supplied to the HWA and the X component of the calorimetric measurement. There is also a strong negative relationship with the quadrature component. These high correlations are expected for a sensor primarily designed to be a thermal anemometer.

In terms of the AoA, the sensor data shows no correlation in terms of P and Y and a weak correlation for X. This in part confirms that the HWA portion (P) of the sensor should be independent of a changing AoA, while the calorimetric portion (X) should be angle-dependent.

The %RH shows a moderately positive correlation with the Y component of the lock-in amplifier. This can be attributed to the contribution of the imaginary component from parasitic capacitances, which causes a phase shift in the measured signal. Furthermore, a lack of correlation to X and P implies little to no thermal interaction with the calorimetric and HWA elements as a result of a change in %RH.

A. ML Models

As Table 3 shows, the LR-X algorithm can only determine the flow velocity. When LR is applied to the entire dataset including P and Y, the flow velocity is predicted more accurately. In addition, values for the other parameters are also estimated. This shows that sampling different portions of the measurement output of the sensor allows for more precise and accurate flow speed measurements. The training time is higher than LR-X due to needing to fit two additional parameters.

Gradient boosting shows similar results compared to LR. The training and inference times are three orders of magnitude larger than those of LR while only having a marginal improvement in accuracy. A reason for its low accuracy compared to RF could be due to RF accounting for multiple features simultaneously, while GB iteratively improves based on one feature at a time [15].

By far the best results in terms of MAE are shown by both the K-NN and RF models. When the data is nonlinear with multiple features these models generally perform better than models, which apply a linear solution or solve using at a single feature [18]. The trained RF model is more complex when compared to the K-NN model due to its large decision-tree-based algorithm and seems slightly more accurate in its results.

The RF's MAE for flow estimation divided by that of the LR-X, gives a 511% improvement. The K-NN model is both faster to train and has the better average inference speed. With a sampling rate of 1 kHz corresponding to 1 ms, measurements take two to three orders of magnitude longer than the inference time of the RF and K-NN models, respectively. Therefore, these models do not create a significant bottleneck when compared to sampling.

In the intended application, one measurement per minute is sufficient while higher accuracy is preferred. As a result, the RF model was chosen to analyze the data in this work despite the speed of the K-NN model. For a standalone embedded system, one could prefer the K-NN model without excessive loss of accuracy due to the smaller model size and lower complexity. Alternatively, it would be worth investigating the effect of reducing the RF models complexity by limiting the number of estimators and encoding the decision tree in float32 or less to reduce memory usage.

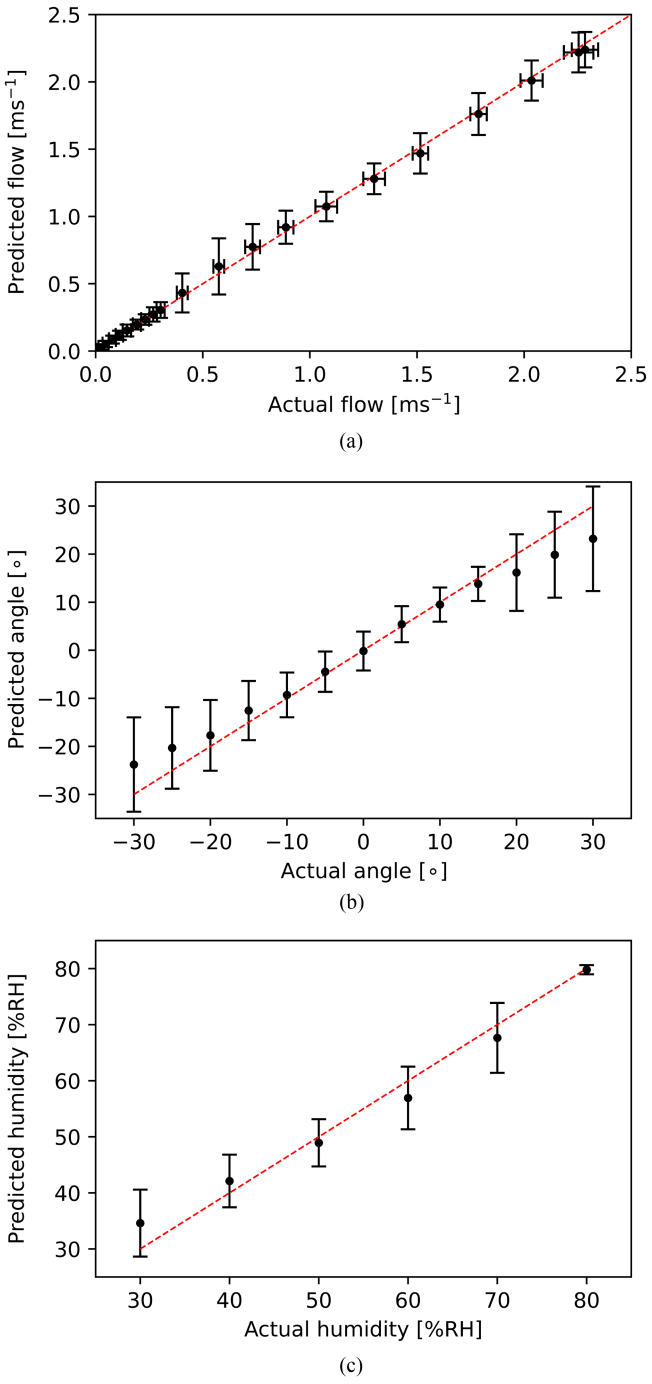


Fig. 3. RF regression predicted results compared to actual results for (a) flow speed, (b) AoA, and (c) %RH. Error bars represent the RMSE value for each prediction. Actual angle and actual humidity have no horizontal error bars due to the errors being so small. The red dashed line represents equality, or a perfect prediction.

B. Flow Speed

Fig. 3(a) shows that the predicted physical parameters are correlated well with the actual values and are along the line of equality. The datapoints represent flow measurements for all possible combinations of flow speed, angle and %RH. In terms of the application in vertical

farms, the resolution of flow speed (<0.05 m s^{-1}) is sufficient and seems to be especially accurate for low flow speeds as shown in the figure. The accurate predictions are due to the high degree of correlation of P, X, and Y with the flow speed comparing well with the state-of-the-art in the range of 0–5 m s^{-1} [3].

C. Angle of Attack

Looking at the prediction of AoA, it becomes clear [see Fig. 3(b)] that the RF model is most accurate between −5° and 15°. Due to the geometry of the calorimetric sensing elements, the measured value of X varies the most around an AoA of 0°, which could explain the increased accuracy. The asymmetry in accuracy at negative angles is due to the introduction of turbulent nonlinearities in flow caused by the protective structural bracket placed around the sensing elements. The RF regression is able to give an indication of the AoA within an MAE of 3.6°, which suffices for the proposed horticultural application.

The reason for the larger errors in predicted angle against the actual angle can be attributed to the angle data having a low correlation with just one output variable (see Table 2). One possible solution would be to separately measure each sensing resistor.

D. Relative Humidity

Fig. 3(c) shows the predicted %RH values compared to the actual humidity of the air in the climate-controlled cabinet. The RF model is able to predict the humidity within an MAE of 2.7%RH. The prediction at 80%RH has a significantly lower root-mean-square error (RMSE) when compared to the rest of the datapoints. This suggests that there exists a different physical interaction between the sensor and the highest humidity. This could be due to high degrees of water adsorption to the sensor surface or other humidity related effects on measurement apparatus. The thermal anemometer has been enhanced through the use of ML and is able to give an estimate of the relative humidity without any additional design considerations.

V. CONCLUSION

By applying ML to the sensor data, the anemometer has improved significantly in both accuracy and functionality. The LR that used the additional power and phase data showed an improved MAE for velocity measurements, but did not discern the additional measurands with enough accuracy for the intended horticultural application. The RF algorithm performed the best for all three measurands and was able to outperform the typical LR by an order of magnitude (511%). Due to the computation speed of RF, no significant sampling rate bottleneck is presented, showing promise for the future development of a ML enhanced embedded system. The use of RF algorithms opens the possibility for multi-purpose sensing with straightforward and cost-effective sensors.

ACKNOWLEDGMENT

This work was supported in part by the Dutch Ministry of Education, Culture, and Science for funding part of the project through Sectorplan Bèta en Techniek.

The authors would like to thank Henk-Willem Veltkamp and Daniël Bijsterveld for aid in fabrication of the anemometers in the MESA+ cleanroom.

REFERENCES

[1] G. Agati, B. Franchetti, F. Rispoli, and P. Venturini, "Thermo-fluid dynamic analysis of the air flow inside an indoor vertical farming system," *Appl. Thermal Eng.*, vol. 235, 2024, Art. no. 121553.

[2] E. Kälvesten, C. Vieider, L. Löfdahl, and G. Stemme, "An integrated pressure-flow sensor for correlation measurements in turbulent gas flows," *Sensors Actuators A: Phys.*, vol. 52, pp. 51–58, 1996.

[3] J. T. W. Kuo, L. Yu, and E. Meng, "Micromachined thermal flow sensors," *Micromachines*, vol. 3, pp. 550–573, 2012.

[4] A. S. Cubukco, E. Zernickel, U. Buerklin, and G. A. Urban, "A 2D thermal flow sensor with sub-mW power consumption," *Sensors Actuators A: Physiccal*, vol. 163, pp. 449–456, 2010.

[5] P. Fürjes, G. Légrádi, C. Dücső, A. Aszódi, and I. Bársony, "Thermal characterisation of a direction dependent flow sensor," *Sensors Actuators A: Phys.*, vol. 115, pp. 417–423, 2004.

[6] S. Hung, S. Wong, and W. Fang, "The development and application of microthermal sensors with a mesh-membrane supporting structure," *Sensors Actuators A: Phys.*, vol. 84, pp. 70–75, 2000.

[7] J. Amaral, J. R. C. Silva, D. S. M. de Andrade, L. T. Ferreira, T. M. Quirino, and J. Quirino, "Machine learning algorithms applied to the inference of the flow rate in a non-intrusive thermal flow meter," in *Proc. 4th Int. Symp. Instrum. Syst., Circuits Transducers*, 2019, pp. 1–6.

[8] W. B. Lenz, U. Yaqoob, R. T. Rocha, and M. I. Younis, "Selective gas detection using condctivity based MEMS resonator and machine learning," in *Proc. IEEE Sensors Conf.*, 2022, pp. 1–4.

[9] F. Mayer, O. Paul, and H. Baltes, "Flip-chip packaging for thermal CMOS anemometers," in *Proc. IEEE 10th Annu. Int. Workshop Micro Electro Mech. Syst.: An Investigation Micro Structures, Sensors, Actuators, Machines Robots*, 1997, pp. 203–208.

[10] C. Liu, L. Du, Z. Zhao, Z. Fang, C. Liu, and L. Li, "A directional anemometer based on MEMS differential pressure sensors," in *Proc. 9th IEEE Int. Conf. Nano/Micro Engineered Mol. Syst.*, 2014, pp. 517–520.

[11] D. Alveringh et al., "A miniature microclimate thermal flow sensor for horticultural applications," in *Proc. IEEE Sensors Conf.*, 2022, pp. 1–4.

[12] T. L. Hackett, D. Alveringh, R. G. P. Sanders, T. E. van den Berg, and J. E. Schmitz, "A multi-parameter measurement system for MEMS anemometers for data collection with machine learning outcomes," in *Proc. 5th MFHS Conf.*, 2024, pp. 126–129.

[13] T. Cover and P. Hart, "Nearest neighbour pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[14] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, 1995, vol. 1, pp. 278–282.

[15] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, pp. 1189–1232, 2001.

[16] F. Pedregosa et al., "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[17] D. S. Soper, "Greed is good: Rapid hyperparameter optimization and model selection using greedy *k*-fold cross validation," *Electronics*, vol. 10, no. 16, pp. 1–23, 1973.

[18] A. Singh, M. N. Halgamuge, and R. Lakshmiganthan, "Impact of different data types on classifier performance of random forest, naïve Bayes, and k-nearest neighbors algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, pp. 1–10, 2017.