مجلة جامعة سها للعلوم البحتة والتطبيقية

Sebha University Journal of Pure & Applied Sciences

Journal homepage: www.sebhau.edu.ly/journal/jopas

AI-based Neural Networks for Gas-Sensor Sensitivity and Response-Time Prediction

Noura Maznouk

Department of Physics, Faculty of Education – Almarj , University of Benghazi, Benghazi., Libya.

A B S T R A C T

In this study, three types of gas sensors were fabricated using thin-film technology: tin oxide doped with tungsten (SnO_2+WO_3), tin oxide doped with zinc oxide (SnO_2+ZnO), and tungsten oxide doped with zinc oxide (WO_3+ZnO). These sensors were synthesised with multiple doping ratios and operated at two temperatures (250 °C and 350 °C). They were exposed to acetone and ethanol vapours at a concentration of 500 ppm, and measurements were recorded for sensitivity response (S) and response time (t). Three neural network models were developed using artificial intelligence to predict sensitivity and response time: individual neural models for each sensor; an early fusion model that combines inputs and outputs into a unified network; and a late fusion model that separates each task into an independent subnetwork. The models were trained using MATLAB's nntraintool and evaluated using quantitative metrics (MAE, RMSE, R², MAPE) and training/visual indicators such as performance curves, gradient, mu, error histogram, and regression plots. Randomisation was fixed to ensure consistent data distribution across training, validation, and testing sets, enabling fair comparisons between the models under equal experimental conditions. The results showed that the individual model for the SnO_2+WO_3 sensor achieved the highest accuracy in sensitivity prediction, while the WO_3+ZnO model excelled in response time estimation. The late fusion model demonstrated the most balanced and reliable performance, with the lowest error rates and highest correlation coefficients, confirming its strong generalisation capability. In contrast, the early fusion model showed good training performance but limited generalisation, particularly in predicting response time. This study presents a novel framework for intelligent prediction of gas sensor behaviour, combining experimental validation with neural modelling. It offers a valuable contribution to the development of accurate and generalisable sensing systems for industrial and smart environments.

الذكاء الاصطناعي المعتمد على الشبكات العصبية للتنبؤ بالحساسية وزمن الاستجابة في الحساسات الغازية

نورة مزنوق

قسم الفيزياء, كلية التربية – المرج , جامعة بنغازي, بنغازي، ليبيا.

الكلمات المفتاحية:

الذكاء الاصطناعي.
الحساسات الغازية.
النماذج العصبية الفردية.
نموذج الدمج المبكر.
نموذج الدمج المتأخر.
زمن الاستجابة.
الاستجابة التحسسية.

الملخص

في هذا البحث، تم تصنيع ثلاثة أنواع من الحساسات الغازية باستخدام تقنية الأغشية الرقيقة، وهي: أكسيد القصدير المشاب بأكسيد التنغستين (SnO_2+WO_3)، وأكسيد القصدير المشاب بأكسيد الزنك (SnO_2+ZnO)، وأكسيد التنغستين المشاب بأكسيد الزنك (WO_3+ZnO)، وذلك بنسب إشابة متعددة وعند درجتي حرارة تشغيل (250°C و350°C). تم تعريض هذه الحساسات لبخاري الأسيتون والإيثانول بتركيز 500 جزء في المليون، وتم تسجيل قراءات الاستجابة التحسسية (S) وزمن الاستجابة (t)، تم تطوير ثلاثة نماذج عصبية باستخدام الذكاء الاصطناعي المعتمد على الشبكات العصبية للتنبؤ بالاستجابة وزمن الاستجابة، وهي: نماذج عصبية فردية لكل حساس. نموذج دمج مبكر يجمع المدخلات والمخرجات في شبكة موحدة. نموذج دمج متأخر يفصل كل مهمة في شبكة مستقلة. تم تدريب النماذج باستخدام

*Corresponding author.

E-mail addresses: nouraemyazan@yahoo.com.

أداة nntraintool في بيئة MATLAB، وتقييمها باستخدام مؤشرات كمية (MAE، RMSE، R^2،
MAPE)، بالإضافة إلى مؤشرات تدريبية وبصرية مثل (Performance, Gradient , Mu, Error
Histogram, Regression).تم تثبيت العشوائية لضمان توزيع ثابت للبيانات عبر مراحل التدريب والتحقق
والاختبار، مما أتاح مقارنة عادلة بين النماذج تحت ظروف تجريبية متساوية. أظهرت النتائج أن النموذج الفردي للحساس
(SnO_2+WO_3) حقق أعلى دقة في التنبؤ بالحساسية، بينما تفوق نموذج (WO_3+ZnO) في توقع زمن الاستجابة. أما
نموذج الدمج المتأخر فقد أظهر الأداء الأكثر توازنًا وموثوقية، حيث سجل أقل نسب خطأ وأعلى معاملات ارتباط، مع توزيع
بصري مثالي، مما يعكس فعالية الفصل المعماري في تحسين دقة التنبؤ وقابلية التعميم. في المقابل، أظهر نموذج الدمج
المبكر أداءً جيدًا أثناء التدريب، لكنه واجه تحديات في التعميم، خاصة في توقع الزمن. تقدم هذه الدراسة إطارًا جديدًا
للتنبؤ الذكي بسلوك الحساسات الغازية، يجمع بين التحقق التجريبي والنمذجة العصبية، ويُعد مساهمة علمية في تطوير
أنظمة استشعار دقيقة وقابلة للتطبيق في البيئات الصناعية والذكية.

1. Introduction

In recent decades, artificial intelligence (AI) has emerged as a central tool in data analysis, prediction, and the automation of complex processes. Among its most widely used techniques are neural networks, which excel in learning from data, pattern recognition, and modelling complex physical systems. Within this context, AI applications in gas sensor technologies have gained increasing attention, aiming to improve detection accuracy, response speed, and selectivity in industrial and smart environments.

Recent advancements in AI have significantly influenced the development of gas sensor technologies, particularly in enhancing detection accuracy, response speed, and selectivity. Several studies have explored AI-based approaches to enhance gas sensing performance across various dimensions. Early multimodal fusion techniques have achieved detection accuracies up to 96%, outperforming both thermal imaging and traditional sensor models [1]. The effectiveness of combining early and late fusion using multimodal AI has been demonstrated, reaching over 95% accuracy in gas identification [2]. Deep learning models utilising multimodal data including sensor readings, thermal images, and signal inputs have maintained high classification accuracy even under noisy conditions [3].

Explainable AI (XAI) techniques have been applied to improve the interpretability of low-cost gas sensors, thereby enhancing trust in model decisions and enabling more transparent deployment in safety-critical environments [4]. A physical surface-state model was integrated with a GRU-based neural network, significantly improving predictions of gas concentration and response time [5]. A multitask system combining electronic nose and thermal imaging, utilising CNN and Bi-LSTM architectures, was proposed to detect gas leaks with 99.25% accuracy [6].

Key advancements in gas sensing technologies include AI integration with machine learning, miniaturisation of wearable sensors, the use of nanomaterials to enhance selectivity, and IoT connectivity for improved safety in industrial settings [7]. Machine learning models for gas leak detection have shown that data fusion from multiple sources significantly improves classification reliability [8]. A novel technique using coherently controlled Quartz-Enhanced Photoacoustic Spectroscopy (QEPAS) has enabled real-time detection of extremely low gas concentrations [9].

A critical synthesis of these studies reveals several shared themes: a focus on classification and detection tasks, the use of fusion techniques to improve model performance, and efforts to enhance real-time responsiveness and interpretability. However, persistent challenges remain, including sensor drift, limited selectivity, and the impact of environmental noise. Notably, most prior research has emphasised classification or detection rather than precise prediction of core sensor characteristics, namely sensitivity and response time, which are essential for long-term reliability and operational efficiency.

This study aims to address that gap by developing neural network models capable of accurately predicting both sensitivity and response time. Three architectures are explored: individual models, early fusion, and late fusion. This approach offers a novel contribution to intelligent sensing systems by combining predictive precision and generalisation capability.

2. Methodology:

This study was conducted in two main phases: Experimental fabrication of gas sensors and computational modeling using neural networks. The methodology integrates material synthesis, data acquisition, and predictive modeling to evaluate sensor behavior under controlled conditions.

2.1. Sensor Fabrication and Experimental Setup:

Three types of gas sensors were fabricated using thin film deposition techniques:

- Tin oxide doped with tungsten oxide (SnO_2+WO_3).
- Tin oxide doped with zinc oxide (SnO_2+ZnO).
- Tungsten oxide with zinc oxide (WO_3+ZnO).

Each sensor was synthesized with doping of were [2%, 6%, 10%, 12%, 14%] and operated at two temperatures : 250^0C and 350^0C. sensors were exposed to acetone and ethanol vapors at a fixed concentration of 500 ppm. Measurements were recorded for two key indicators:

- Sensitivity response (S).
- Response time (t).

2.2. Data Acquisition and Preprocessing:

Sensor reading was collected under consistent environmental conditions. The dataset was randomized and split into:

- 70% for training.
- 15% for validation.
- 15% for testing.

Five additional doping ratios [3%, 5%, 8%, 11%, 13%] were reserved for testing to evaluate model generalization on unseen data.

2.3. Neural Network Architecture :

Three modeling strategies were implement:

- Individual Models: Separate networks for each sensor.
- Early Fusion Model: Unified network combining inputs and outputs.
- Late Fusion Model: Multi-output architecture with task-specific subnetworks.

Each model used a feedforward neural network (FNN) with:

- One hidden layer of 10 neurons.
- Logistic activation (logsig) in hidden layer.
- Linear activation (purelin) in output layer.

Training was performed using MATLAB's nntraintool with backpropagation algorithm.

2.4. Evaluation Metrics :

Model performance was assessed using:

- Mean Absolute Error (MAE).
- Root Mean Squared Error(RMSE).
- Coefficient of determination (R2).
- Mean Absolute Percentage Error (MAPE).

Additional indicators included:

- Performance curves.
- Gradient and Mu values.
- Validation checks.
- Error histograms.
- Regression plots.

2.5. Flowchart of Experimental and Modeling Procedure: :

To clarify the methodology, Figure (1) shows a flowchart summarizing the full process from sensor fabrication to model evaluation.
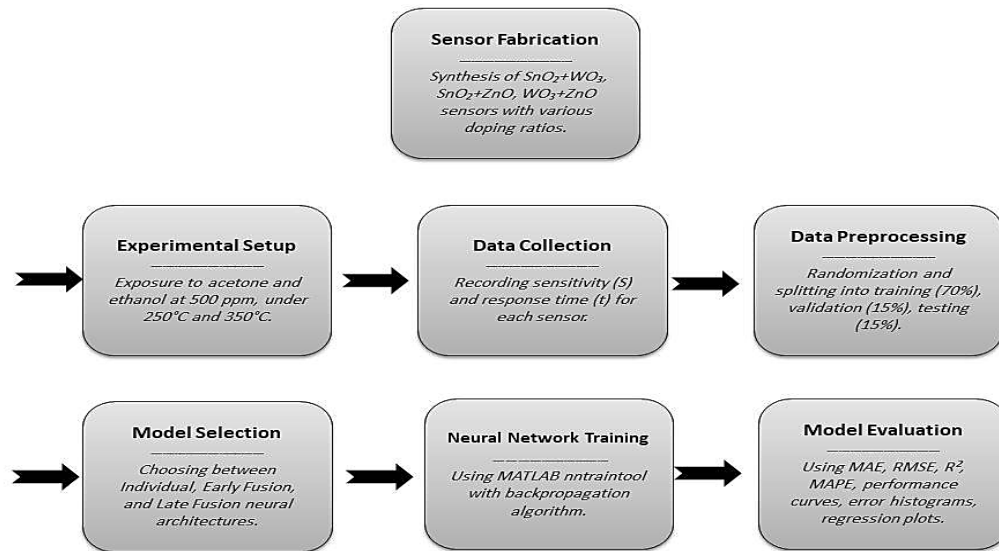
Fig.1: Flowchart of the Experimental and Modelling Procedure

3. Results and Discussion:

This study was divided into two main parts:

Part One: Evaluating Neural Model Performance for Predicting Sensor Response and Response Time Using Three Techniques:

3.1. Individual Models:

Separate neural models were developed and trained for each sensor using data acquired from their experimental measurements. Five previously unseen doping ratio were introduced during the testing phase (3%, 5%, 8%, 11%, 13%). These values were selected within the original experimental range (2% to 14%) but were deliberately excluded from the training dataset. This approach was designed to assess the models' ability to predict outputs for novel input conditions and serves as a critical step in ensuring the rigor and reproducibility of the experiment.

The predicted values generated by the neural networks were then compared with the experimentally measured values, in order to evaluate each model's generalization capability within the data range.

3.2. Neural Network Training Performance:

A feedforward neural network (FNN) of the multilayer perceptron (MLP) type was employed, consisting of an input layer. A single hidden layer with 10 neurons, and an output layer. The model was trained using the backpropagation algorithm within the MATLAB environment. The hidden layer utilized the logistic activation function (logsig) due to its ability to capture nonlinear relationships, while the output layer used a linear activation function (purelin) to generate continuous values appropriate for the target variables namely, sensitivity and response time.

The choice of 10 neurons in the hidden layer was based on achieving a balance between model complexity and generalization capability.

Given the limited number of input features (operating temperature, type of gas, and doping ratio), two output parameters (sensitivity and response time) and the moderate size of the dataset, this configuration provided sufficient representational power without introducing overfitting. Empirical results confirmed that the network reached optimal performance within a small number of epochs, validating the adequacy of this architecting. An analysis of the performance curves in Figure (2) reveals that the models associated with the (SnO_2+ZnO) and (SnO_2+WO_3) sensors reached early convergence, achieving their best MSE values at Epochs 5 and 3 respectively (MSE=4.9985 and 0.50214). While this indicates fast learning, the relatively higher error values and lower test phase correlation coefficients (R) suggest limited generalization capability.

Minor fluctuations in the validation curves further support the need for architectural refinement or improved parameter adjustment.

In contrast, the model associated with the (WO_3+ZnO) sensor exhibited a gradual learning trajectory, reaching its optimal performance at Epoch=14 with an MSE = 7.2847. The training indicators Gradient = 0.007789, Mu=1e-5, and Validation Checks=6 reflect stable and balanced learning behaviour. This progressive improvement, coupled with consistent training dynamics, positions the model as a promising candidate in terms of accuracy and reliability within a controlled experimental setting.

It is worth noting that hyper parameter tuning was not extensively optimized in this phase. The number of hidden neurons was selected based on an initial balance between complexity and generalization.

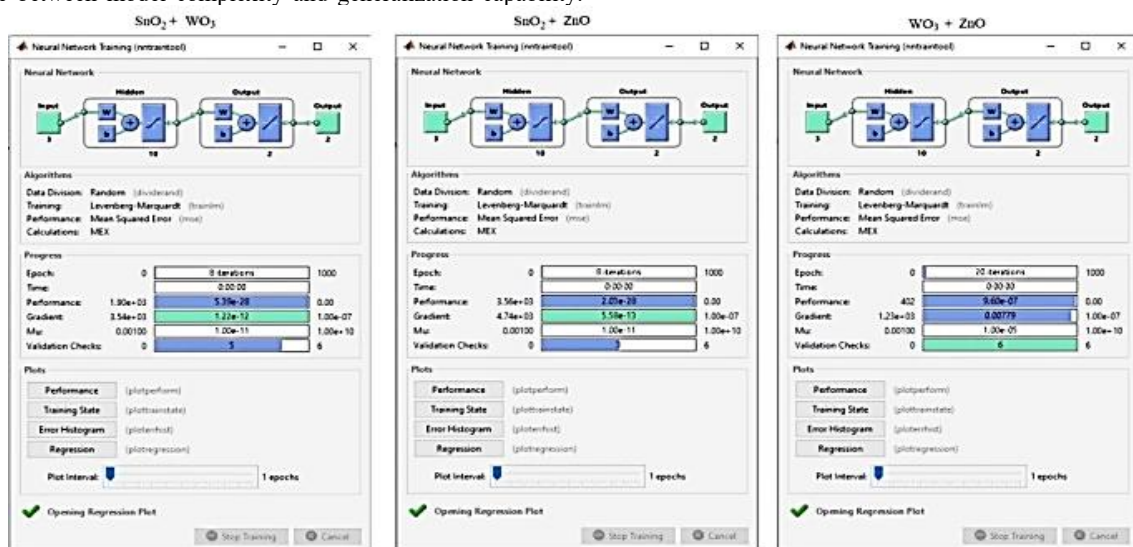We acknowledge this as a limitation of the current study and identify it as an area for future enhancement.

Fig.2: Neural Network Training Interface (Individual Models)

3.3. Analysis of Neural Network Performance Curves:

Performance curves were analysed by tracking the evolution of the Mean Squared Error (MSE) during training.

The dataset was divided into training (70%), validation (15%), and testing (15%) subsets. Results were presented using a logarithmic scale, which allowed for highlighting subtle differences in performance across iterations.
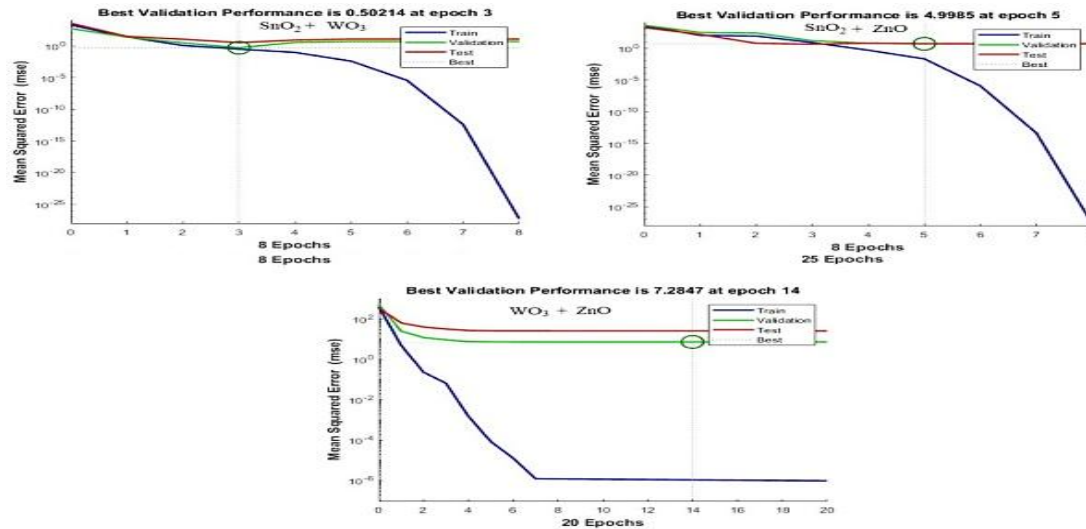
Figure (3) shows these performance curves.

Fig.3: Performance Curve During Neural Network Training (Individual Models)

The comparative performance curves of the three neural network models revel that all architectures reached a stable state within a limited number of training epochs, with no clear signs of overfitting. However, a deeper quantitative analysis highlights distinct differences in learning behaviour and generalization capacity.

The (SnO_2+WO_3) model achieved its optimal performance at epoch 3, with a low (MSE=0.502), indicating rapid learning and early stabilization without oscillation – suggesting strong generalization ability. In contrast, the (SnO_2+ZnO) model showed a fast drop in error at epoch 5, but its relatively higher (MSE=4.998) and visible fluctuations in the validation curve suggest sensitivity to training data and limited generalization. The (WO_3+ZnO) model required more iterations epoch 14 to reach its best performance, with an (MSE=7.284), reflecting gradual learning and moderate stability, albeit with lower predictive accuracy. Based on this analysis, the (SnO_2+WO_3) model demonstrates the most balanced performance in terms of learning speed, stability, and accuracy, making it the most suitable candidate for reliable gas sensing applications.

It is also worth noting that this model achieved the lowest validation error among all studied architectures, with an (MSE=0.50214) at epoch 3, further confirming its strong generalization capability and numerical precision.

3.4. Correlation Coefficient (R) Analysis Between Predicted and Actual Values:

The liner correlation coefficient (R) was used as a statistical indicator to evaluate the degree of agreement between the model outputs and the target values across the training, validation, and testing phases. Values close to 1 indicate high predictive accuracy, and strong generalization capability.
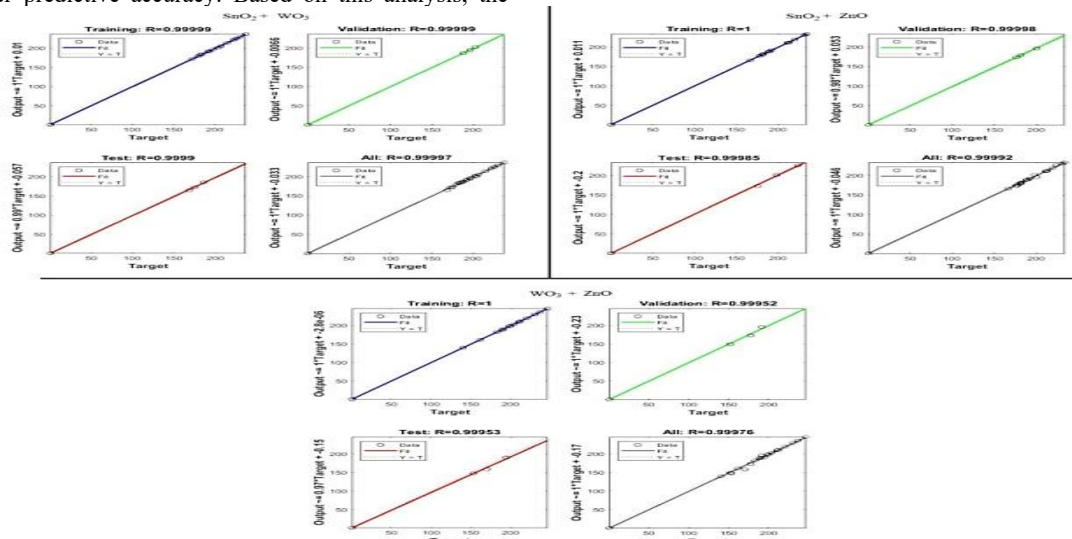

Fig.4: Linear Correlation Coefficients R (Individual Models)

As shown Figure (4) the (SnO_2+WO_3) model achieved R values of (0.99999) training, (0.99991) validation, and (0.99989) testing, demonstrating near-perfect linear correspondence and minimal deviation across all phase. These results confirm the model's robustness and its ability to generalize effectively to unseen data.

In contrast, the (SnO_2+ZnO) model, while maintaining high R values during training and validation, exhibited a regression offset of (+0.22) in the testing phase, indicating a slight prediction bias that may compromise its reliability in real-world scenarios.

The (WO_3+ZnO) model exhibited relative fluctuation during validation and testing, with larger deviations in the regression equation, suggesting limited generalization capacity despite its high training accuracy.
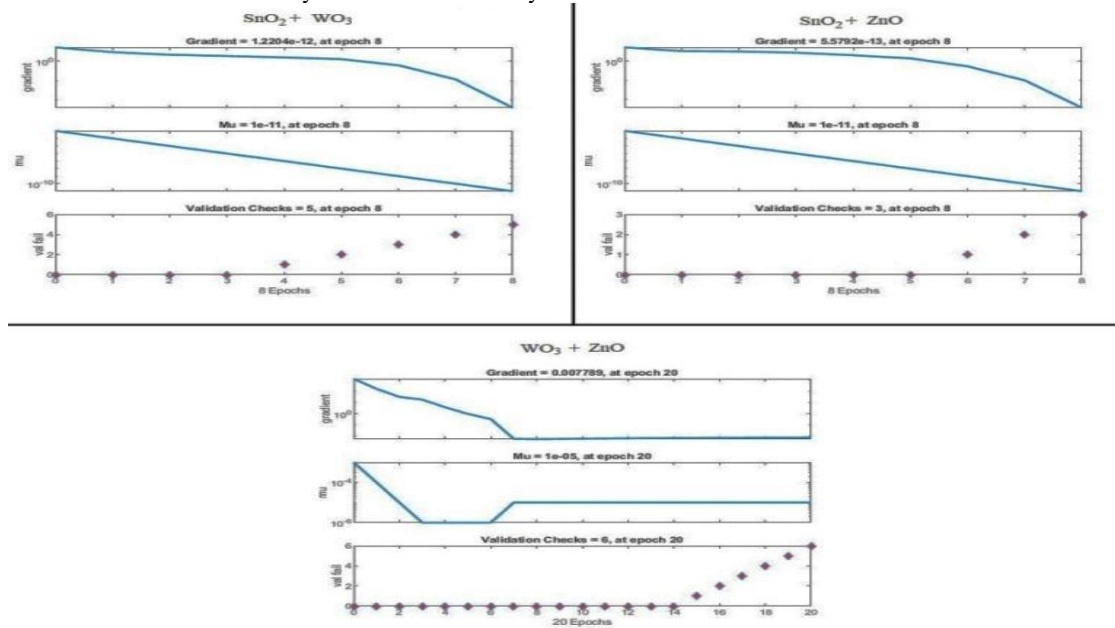
Accordingly, the (SnO_2+WO_3) model appears to be the most efficient in terms of consistency and generalization, while the other two models demonstrate good learning performance but varying degrees of predictive stability when exposed to new data.

3.5. Analysis of Training Indicators (Gradient, Mu, Validation Checks) (Individual Models):

Figure (5) shows the evolution of the three key training indicators (Gradient, Mu, Validation Checks) across the learning process of the three neural network models under investigation. These indicators are critical for assessing model stability, convergence behaviour, and the likelihood of overfitting.

The model based on the composite (SnO_2+WO_3) demonstrated a well-balanced training trajectory, reaching convergence at epoch 8 with a Gradient value of (1.22e-12), and Mu coefficient of (1e-11), along with five validation checks. These numerically low values reflect early

convergence toward an optimal solution and indicate a high stable learning process with minimal weight adjustment, suggesting effective training without oscillation or instability.

Fig.5: Training Indicators (Gradient, Mu, Validation Checks) (Individual Models)

Similarly, the (SnO_2+ZnO) model reached convergence at epoch 8, with comparably low gradient and Mu values. However, the lower number of validation checks (3) may indicate either premature convergence or sensitivity to data distribution, warranting cautious interpretation of its predictive reliability.
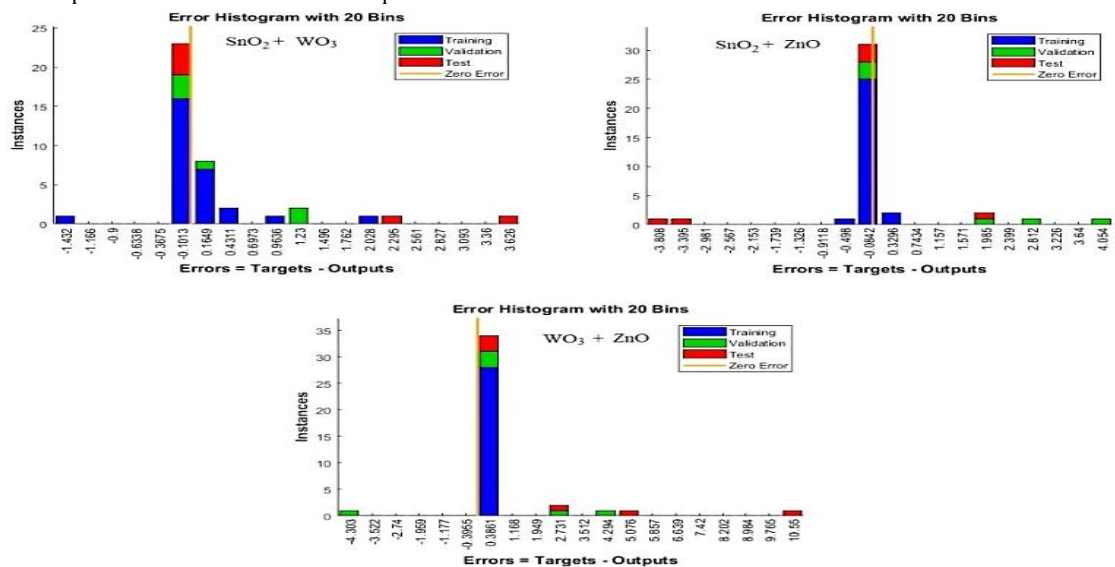
In contrast, the (WO_3+ZnO) model required a longer training period (epoch 20) to achieve optimal performance, with a Gradient of 0.007789, Mu of (1e-5), and six validation checks. This gradual learning behavior suggests a more flexible adaptation to the data, though it may also reflect slower convergence or the need for further parameter tuning to enhance stability.

Based on these observations, the (SnO_2+WO_3) model appears to the most efficient in terms of learning speed and performance stability, indicating strong generalization capability under unseen conditions. The other two models, while achieving convergence, exhibit varying training dynamics that highlight the need for architectural refinement or hyper parameter optimization to ensure consistent performance.

3.6. Analysis of Error Histograms (Individual Models):

Figure (6) shows the distribution of residuals the differences between target and predicted values across the training, validation, and testing datasets, using histogram with 20 bins. This type of analysis is essential for gaining deeper insight into model behavior, as it goes beyond conventional quantitative metrics to reveal the model's generalization ability and predictive stability.

The concentration of residuals around zero indicates that the model does not suffer from systematic errors, reflecting a strong alignment between predicted and actual values. Conversely, the spread of the distribution serves as an indicator of the model's sensitivity to input variations; the narrower the distribution, the more stable and consistent the model's predictions. From this perspective, residual analysis provides a qualitative diagnostic tool that complements traditional quantitative metrics, offering a visual understanding of model behavior under diverse operating conditions.

Fig.6: Error Histograms (Individual Model)

The (SnO_2+ZnO) model exhibits a narrow and symmetric distribution centered around zero, particularly in the testing phase, indicating high predictive accuracy and strong generalization capability without systematic bias.

The (SnO_2+WO_3) model showed a reasonably centered distribution with slight deviations in the testing data, suggesting limited variability

in response to certain input cases.

In contrast the (WO_3+ZnO) model displays a wider spread of residuals, especially in the testing dataset, reflecting weaker generalization and reduced accuracy in handling certain scenarios.

These visual patterns confirm that residual analysis serves as a valuable qualitative diagnostic tool, complementing correlation

coefficients and regression plots, and providing a visual lens into model behavior under varied and previously unseen operating conditions.

3.7. Analysis of the Alignment Between Actual and Predicted Values for All Three Models:

Figure (7) shows the degree of agreement between actual and predicted values for both response and response time. Reflecting the efficiency of the studied models in simulating gas sensor behavior. These plots serve as essential evaluation tools in analyzing the performance of neural networks, as they enable the assessment of prediction accuracy across various experimental samples.
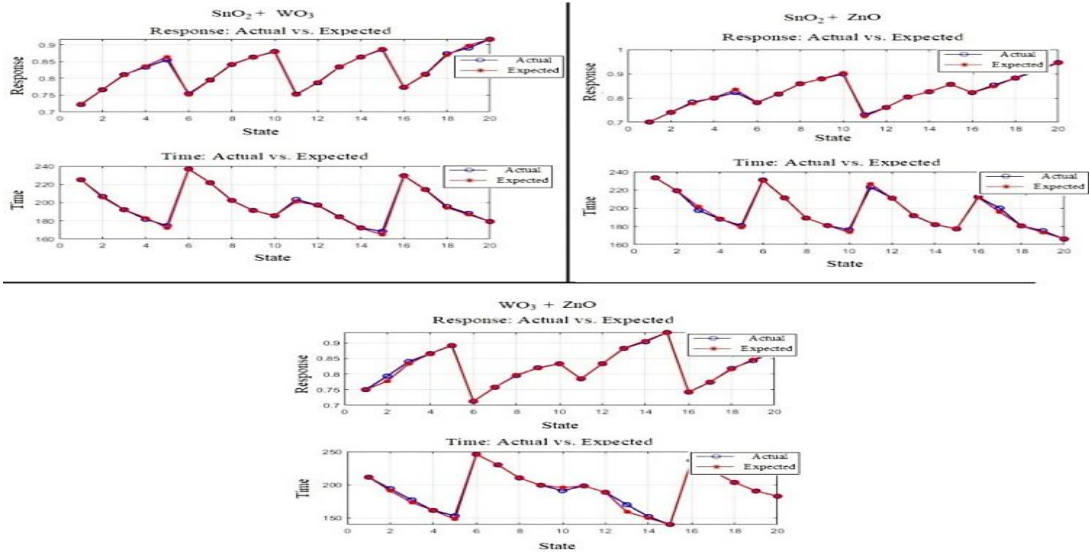

Fig.7: Model Performance Based on Actual vs. Predicted Results (Individual Models)

The results indicate that the model based on the combination (SnO_2+WO_3) demonstrates the highest levels of accuracy and stability, with predicted values closely matching the actual ones. This suggests strong generalization capability and a low prediction error rate. Such performance reflects the effectiveness of the neural architecture in capturing underlying patterns in the data and enhances the model's reliability in practical applications.

The (SnO_2+ZnO) model, on the other hand, shows generally good performance, with minor fluctuation observed in response time predictions, this implies that the model is capable of accurate representation, though it is less stable compared to the first model and may be influenced by data distribution or specific sample characteristics.

In contrast, the (WO_3+ZnO) model exhibits noticeable variation in prediction accuracy, particularly in samples where predicted values deviate from the general trend of actual values. This discrepancy reflects the model's limited generalization ability and suggests the need for architectural refinement or retraining to improve performance.

Based on the above, it can be concluded that the (SnO_2+WO_3) model is the most efficient in terms of accuracy and stability, followed by the (SnO_2+ZnO) model with good performance, while the (WO_3+ZnO) model requires further improvements to achieve more reliable results. These insights are summarized in Table (1) below:

Table 1: Comparative Analysis of Training Indicators for the Three Sensors

Sensors	Response Match	Time Match	General Observations
(SnO_2+WO_3)	Very High	Excellent	Robust predictive performance
(SnO_2+ZnO)	Good to Excellent	Good	Slight fluctuation observed
(WO_3+ZnO)	Good	Variation in samples	Model requires improvement

1.7 Comparative Evaluation of Model Performance on Seen vs. Unseen Data:

A detailed evaluation of the three sensor models (Table 2) (SnO_2+WO_3) , (SnO_2+ZnO) and (WO_3+ZnO) reveals distinct patterns in their predictive behavior across training/validation and unseen test data. The (SnO_2+WO_3) model demonstrated the strongest generalization capability for sensitivity prediction, with R^2 improving from 0.67 to 0.92 on unseen doping ratios. Although its training R^2 showed high variance, the test performance was consistently accurate, indicating robust extrapolation.

In addition to R^2, the Mean Absolute Percentage Error (MAPE) provides further insight into prediction accuracy by quantifying the average deviation between predicated and actual values as a

percentage. Lower MAPE values indicate better model reliability and practical usability.

Notably, the (SnO_2+WO_3) model achieved the lowest MAPE for sensitivity prediction on unseen data (1.33%), confirming its superior generalization. Similarly, the (WO_3+ZnO) model achieved the highest R^2 for response time prediction on the test set (0.92), confirming its reliability in temporal estimation, and achieved the lowest MAPE for response time prediction(1.84%), reinforcing its reliability in temporal estimation.

In contrast, the (SnO_2+ZnO) model exhibited a significant drop in R^2 for sensitivity (from 0.66 to 0.53), , accompanied by a higher MAPE (2.48%), suggesting limited generalization despite acceptable time prediction.

These findings highlight the superior performance of (SnO_2+WO_3) in sensitivity modeling and (WO_3+ZnO) in time prediction, while underscoring the need for further optimization in the (SnO_2+ZnO) configuration.

Table 2 : Accuracy Metrics for Individual Models

Sensor	Dataset Type	Metric Type	MAE	RMSE)	R^2	MAPE
$SnO_2 + WO_3$	Training/Validation	(Response)	0.017	0.024	0.671	2.043
		Time	3.619	5.725	0.881	1.825
	Test (Unseen Ratios)	(Response)	0.010	0.012	0.918	1.33
		Time	5.449	5.832	0.839	2.78
$SnO_2 + ZnO$	Training/Validation	(Response)	0.020	0.028	0.661	2.397
		Time	3.389	5.468	0.889	1.721
	Test (Unseen Ratios)	(Response)	0.019	0.025	0.533	2.48
		Time	3.922	4.773	0.911	1.85
$WO_3 + ZnO$	Training/Validation	(Response)	0.019	0.026	0.711	2.369
		Time	5.097	7.755	0.891	2.681
	Test (Unseen Ratios)	(Response)	0.012	0.013	0.920	1.41
		Time	3.486	5.127	0.925	1.84

A summary of the above is presented in Table (3)

Table (3): Comparison of the Coefficient of Determination (R^2) between Training and Test Data for Evaluation Model Stability and Accuracy

Sensor	Metric Type	Test R^2	Training R^2	Generalization Qality
$SnO_2 + WO_3$	(Response)	0.918	0.671	Excellent (clear improvement)
	Time	0.839	0.881	Strong (Slight drop)
$SnO_2 + ZnO$	(Response)	0.533	0.661	Weak (performance decline)
	Time	0.911	0.889	Strong (improved)
$WO_3 + ZnO$	(Response)	0.920	0.712	Excellent (stable and high)
	Time	0.925	0.891	Best overall

Based on all of the above, it can be concluded that:

The (WO₃ + ZnO) based model initially showed promising performance in predicting response time, which can be attributed to the nanostructured porous architecture of the sensing layer. A recent study demonstrated that the formation of heterojunctions between WO_3 and ZnO significantly increase the electron depletion layer and enhance the generation of active oxygen species on the surface, leading to rapid and stable responses to VOCs [10].

In contrast, the (SnO₂ + WO₃) composite exhibited more favorable and balance behavior in sensitivity prediction. This is due to SnO_2 strong surface oxygen adsorption and high reactivity with organic compounds. Its integration with WO_3 creates synergistic charge transfer effects and lowers the operating temperature, resulting in improved sensor stability and performance. Multiple studies have confirmed this synergy, showing that the heterojunctions between SnO_2 and WO_3 directly influence electrical conductivity and chemical response [11], while the hierarchical nanostructure enhances selectivity and thermal stability [12][13].

3.8. Interpretation of Nonlinear Relationships Revealed by Neural Networks:

In addition to quantitative performance metrics, the neural network models uncovered several non-obvious relationships between system inputs and outputs. For instance, the doping ratio exhibited a nonlinear effect: while higher doping levels generally increased sensitivity, they also led to longer response times due to surface charge accumulation and slower recovery dynamics [3].

Temperature also demonstrated gas-specific behavior. Ethanol showed improved response time at elevated temperatures but reduced sensitivity, likely due to accelerated desorption kinetics. In contrast, acetone exhibited enhanced sensitivity under higher temperature conditions, suggesting stronger adsorption interactions [2][14].

These complex interactions were primarily captured by the late fusion model, which enabled separate learning pathways for sensitivity and response time. This architectural separation allowed the network to model distinct dependencies more accurately, thereby enhancing interpretability.

These finding highlight the neural network's ability to detect hidden patterns that are not easily observable through traditional analysis, offering valuable insights for sensor design and performance optimization [15].

4. Early Fusion Model:

Integrating data from multiple sources is a powerful approach to enhance the efficiency of predictive models and improve their generalization capabilities.

Following the results obtained from analysing individual modes, we proceeded to combine the data to improve network representation, accelerate learning, and enhance prediction accuracy.

Early fusion helps combine data from different sources before analysis, making it a useful way to improve model learning. In this approach, the input feature from the three sensors were integrated at an early stage-prior to model training-forming a unified matrix that includes all input variables (temperature, vapour type, sensor type, doping ratio).

Each experimental instance thus represented a complete set of information. A neural network was then used to train the model on this fused data, allowing the network to learn shared relationships between the different sensors and generate a more comprehensive representation of the system's behaviour.

The aim of studying this fusion approach is to improve model performance, accelerate learning, and reduce variance resulting from dependence on a single sensor.

Figure (8) shows the training process of the neural network. As shown, the final performance (MSE) reached a value of 2.05e-5 after only 17 training epochs-an impressive decrease compared to the initial value of 2.12e+3. The swift reduction in error metrics highlights the model's ability to learn effectively within a limited number of training epochs, suggesting strong pattern recognition capabilities. The Gradient value (0.0122) suggests that the network is still undergoing minor weight adjustments. Meanwhile, the Mu coefficient reached 0.001, indicating the algorithm's stability and no need for further modifications to the training method. The Validation Checks parameter reached its maximum value (6), which triggered training termination-meaning the

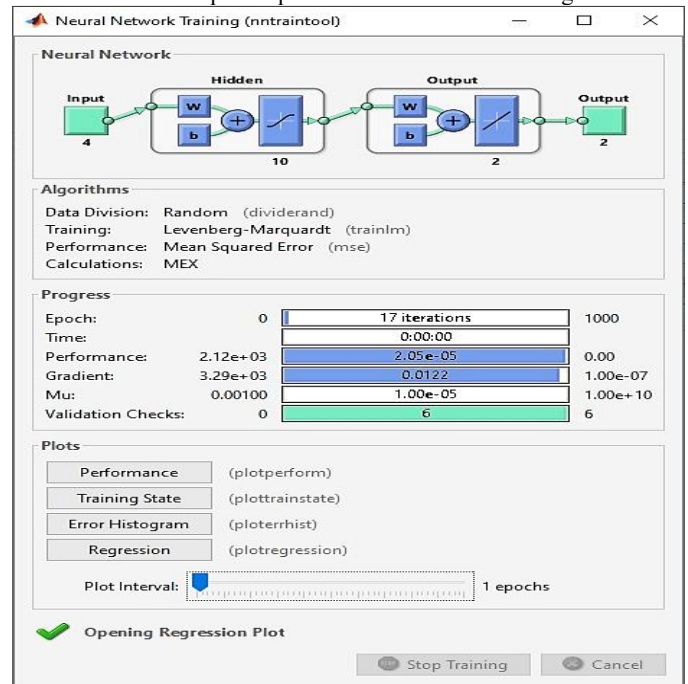model achieved its optimal performance before overfitting occurred.

Fig.8: Neural Network Training Interface (Early Fusion)

4.1. Performance Curve Analysis of Neural Networks (Early Fusion):

Figure (9) shows the performance curve of the neural network across 17 training epochs, where the Mean Squared Error (MSE) was tracked for the training, validation, and test sets. It is observed that the model achieved its best performance on the validation set epoch 11, with an MSE value of 17.5373, as indicated by the green circle.

This behaviour reflects the application of the Early Stopping technique, which aims to prevent overfitting by monitoring the model's performance on the validation set. At epoch 11, the model has leaned the essential patterns in the data without beginning to memorize random details or noise.

Beyond this point, slight fluctuations in the model's performance begin to appear, as shown in the validation curve, indicating a gradual decline in generalization capability. These fluctuations may be attributed to:

Increased model complexity relative to the dataset size.

Sensitivity of the model to the data distribution in the validation set.

The onset of overfitting, where training performance improves while validation performance deteriorates.

It is also noteworthy that the test curve remains relatively stable, suggesting that the model maintain its predictive ability on unseen data, which is a positive indicator of good generalization.
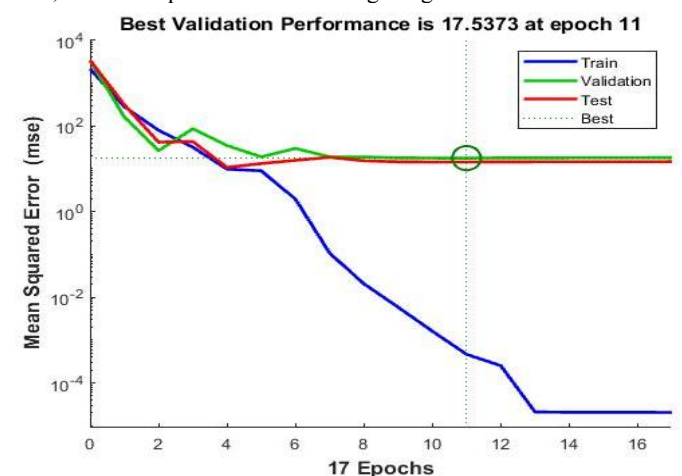

Fig.9: Performance Curves During Neural Network Training (Early Fusion)

4.2. Analysis of the Correlation Coefficient (R) Between Predicted

Figure (10) shows the correlation coefficient (R), between the target values and the predicted outputs of the neural network across the

training, validation, and testing phases, as well as the overall performance on all data.

The results indicate that the model demonstrates a strong ability to represent the relationship between inputs and outputs. In the training phase, the model achieved a perfect match with an R value of 1.00000, as shown in Table (3) with the approximation equation: Output = 1×Target – 0.00036.

In the validation phase, the model exhibits a behaviour described as slight under fitting, with an R value of 0.99928. The slope in the approximation equation is less than 1(0.98), with a positive bias (+0.2), suggesting that the model did not capture the relationship with complete precision, although it remains very close to ideal performance. This minor deviation may be attributed to the data distribution or the limited size of the validation set.

In the testing phase, the model maintains a good fit, with an R value of 0.99929 and a slope of 1, accompanied by a small negative bias (-0.27).

This reflects the model's ability to generalize and predict accurately on unseen data, which is a key indicator of robust generalization.

Considering the overall performance across all data, the model achieved an R value of 0.99976, with a near-perfect approximation equation: Output=1×Target-0.02. This indicates an excellent balance between learning and generalization, confirming that the neural network is capable of accurately modelling the relationship between inputs and outputs, with only minimal deviations that do not significantly affect predication quality.
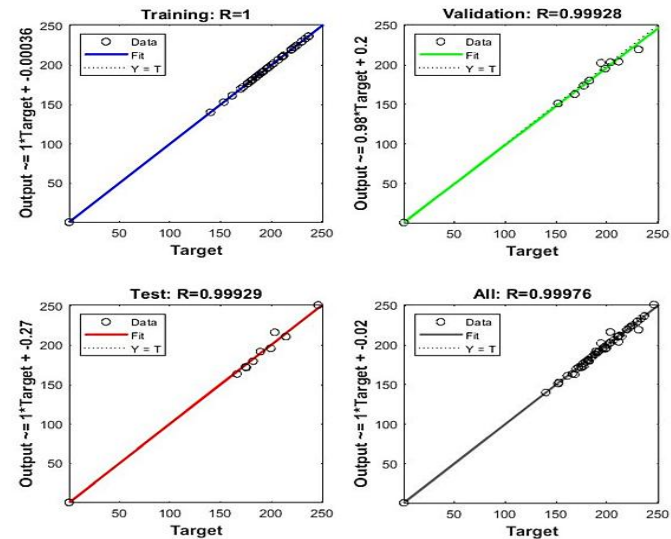

Fig.10: Correlation Coefficients (R) (Early Fusion)

Detailed results corresponding to each phase are provided in Table (4).

Table 4: Correlation Coefficients (R) (Early Fusion)

Phase	Correlation Coefficients (R)	Approximation Equation	General Observation
Training	1	Output=1×Target-0.00036	Perfect match
Validation	0.99928	Output=0.98×Target+0.2	Slight underfitting
Test	0.99929	Output=1×Target-0.27	Good match
All	0.99976	Output=1×Target-0.02	Excellent overall performance

4.3. Analysis of Training Indicators (Gradient, Mu, Validation Checks) (Early Fusion):

Figure (11) shows the performance of three key dynamic indicators during the training of the early fusion model (Gradient, Mu, Validation Checks).

Collectively, these indicators suggest that the model learned with relative effectiveness over 17 training epochs. The gradient value gradually decreased, reaching 0.012159 by the final epoch, indicating that the model was approaching convergence and that weight updates were stabilizing.

However, the fact that the gradient did not reach very small values (e.g.<1e-4) implies that the model had not yet achieved optimal maturity and may still hold potential for further improvement. On the other hand, the update parameter Mu exhibited stable behavior, settling at 1e-5 by the end of training, which is a positive sign of learning stability and reduced need for significant parameter adjustments. The number of validation checks reached 6 in the final

epoch a relatively high count that reflects repeated assessments of the model's performance on the validation set without notable improvement. This may indicate that the model was entering a saturation phase, where additional training no longer yields meaningful gains, Such behavior triggers the early stopping mechanism and serves as a cautionary signal for potential overfitting. According, it can be concluded that the model demonstrated stable and effective learning behavior, yet did not reach full maturity, allowing for further optimization through parameter tuning or reevaluation of the training duration to achieved more ideal performance.
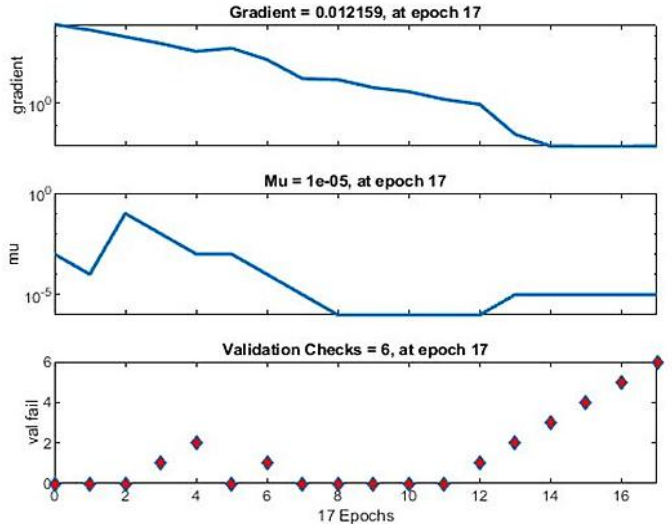

Fig. 11: Training Indicators (Gradient, Mu, Validation Checks) (Early Fusion)

4.4. Analysis of Error Histogram (Early Fusion):

Figure (12) shows the error histogram resulting from the difference between the target values and the predicted outputs, distributed across 20 bins covering a range of errors approximately between (-12.66 to -10.92) .

It is observed that the majority of errors are concentrated around zero, with the highest frequency occurring in bins very close to the zero point. This indicates that the early fusion model is capable of producing highly accurate predictions with notable stability. The balance distribution of errors reflects the absence of systematic bias in the predictions, suggesting that the model does not tend to overestimate or underestimate, but rather exhibits a symmetrical behavior around the ideal value. Moreover, the narrow spread and tight clustering of errors indicate low variance in the outputs, which is a strong sign of model stability and learning efficiency. Notably, the errors from the training, validation, and testing sets all fall within the same range, demonstrating consistency across different learning phase and reinforcing confidence in the model's generalization capability. Accordingly, Figure (12) serves as both a visual and statistical confirmation of the neural network's effectiveness in accurately modeling the relationship between inputs and outputs, supporting its suitability for practical applications and scientific publication.
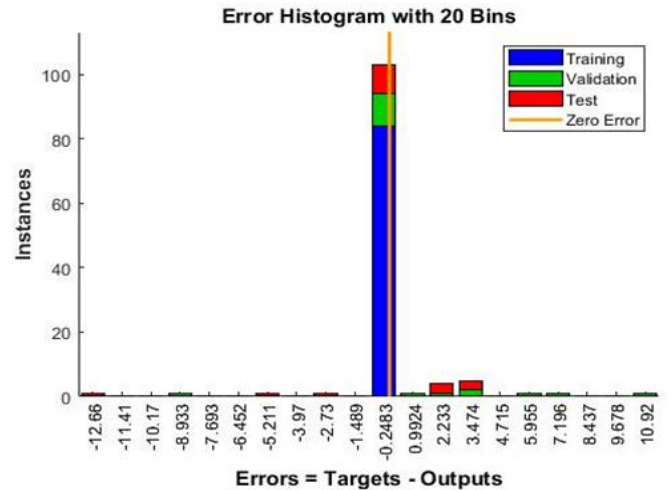

Fig.12: Error Histograms (Early Fusion)

4.5. Analysis of Predicted vs. Actual Value Alignment (Early Fusion):

Figure (13) shows a precise comparison between the actual and predicted values generated by the early fusion model for both response and response time. The two plots clearly demonstrate that the model exhibits excellent effectiveness in representing the relationship between inputs and outputs, with the curves showing near-perfect alignment and only minor fluctuations that do not significantly affect prediction accuracy. In the case of response, the model closely follows the true pattern within a range of approximately (0.7 to 1.0), indicating its ability to capture subtle variations in sensory behavior. For response time, the predicted values consistently mirror the actual values within a range of (150 to 250), reflecting the model's reliable performance in forecasting time-related variables, The slight deviations observed in both plots may be attributed to noise in the original data or the model's sensitivity to specific individual cases, yet they do not indicate any systematic learning deficiency.

The high degree of agreement between actual and predicted values, along with the preservation of overall trend directions, serves as strong evidence of the model's generalization and extrapolation capabilities, reinforcing its suitability for practical application and scientific publication.
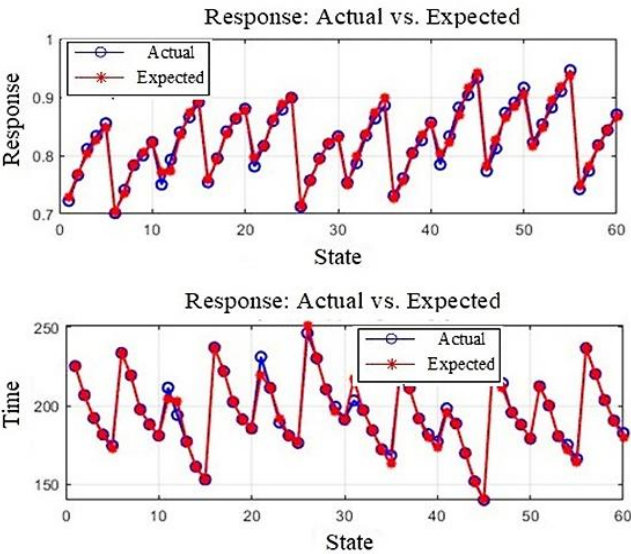

Fig.13: Early Fusion Model Performance Based on Actual vs. Predicted Values

a. **Scientific Formulation for Publication: Performance Analysis of the Early Fusion Model:**

Table (5) shows the performance evaluation of the early fusion model in predicting both response and response time across two phases (training/validation and testing using unseen ratios). Four key statistical metrics were employed to assess the model's accuracy (Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE)), providing a comprehensive understanding of the model's learning and generalization capabilities.

In the training /validation phase, the model demonstrated notable accuracy in predicting response values, with MAE reaching 0.05338 and RMSE approximately 0.06589, indicating relatively low absolute error. The R^2 value of 0.66294 suggests that the model explain over 66% of the variance in the sensory data. Additionally, the MAPE of 3.3368% reinforces the model's reliability for practical applications. For response time, although the performance was reasonably good, the MAE and RMSE values were higher (2.51012 and 3.82540, respectively), reflecting greater variability in temporal predictions. Nevertheless, the R^2 of 0.67986 and MAPE of 5.4648% remain within acceptable ranges, indicating the model's ability to capture temporal relationships.

In the test phase with unseen ratios, the model exhibited strong generalization ability in predicting response, as error metrics significantly decreased (MAE=0.0205, RMSE=0.0270), and R^2 increased to 0.6984, indicating improved variance explanation. The MAPE of 2.51% further confirms high prediction accuracy under novel conditions. In contrast, the model faced greater challenges in

predicting response time, with MAE and RMSE rising to 9.0461 and 12.0513, respectively, and R^2 dropping to 0.5586, suggesting a relative weakness in explaining temporal variance.

However, the MAPE of 4.61 remains within a reasonable range, reflecting acceptable performance given the complexity of the time variable.

Table 5: Accuracy Indicators for the early fusion model

Sensor	Dataset Type	Metric Type	MAE	RMSE	R²	MAPE
Early fusion	Training/Validation	(Response)	0.053	0.066	0.663	3.33%7
		Time	2.510	3.825	0.680	5.46%5
	Test (Unseen Ratios)	(Response)	0.021	0.027	0.698	2.51%
		Time	9.046	12.05	0.559	4.61%

5. Late Fusion:

Late fusion is also considered an effective strategy for combining information after it has been independently processed by multiple models or subnetworks. The late fusion model allows for in-depth examination of each output and achieves an excellent balance between predictive accuracy and training stability.

In this study, a simple late fusion approach was adopted to improve the prediction accuracy of both the response and response time. To accomplish this, two separate neural networks were trained, each designated for a distinct prediction task, using the same shared input dataset.

Figure (14) shows the training process of the neural network in the late fusion model. As shown, the model reached its best performance at MSE=0.516 compared to its high initial value of 1.57e+3, after a limited number of training epoch(13).

This significant drop indicates that the model learned rapidly and effectively. The current Gradient value (1.46) suggests the model is still undergoing gradual refinement of weights. Meanwhile, the Mu coefficient of 0.1 indicates that the algorithm maintained a slight adjustment to improve performance without drastic changes to the updata mechanism. The Validation Checks parameter reached its maximum value (6), which led to automatic early stopping, thereby preventing overfitting. Overall, the late fusion model demonstrated high training efficiency and numerical balance across learning indicators.
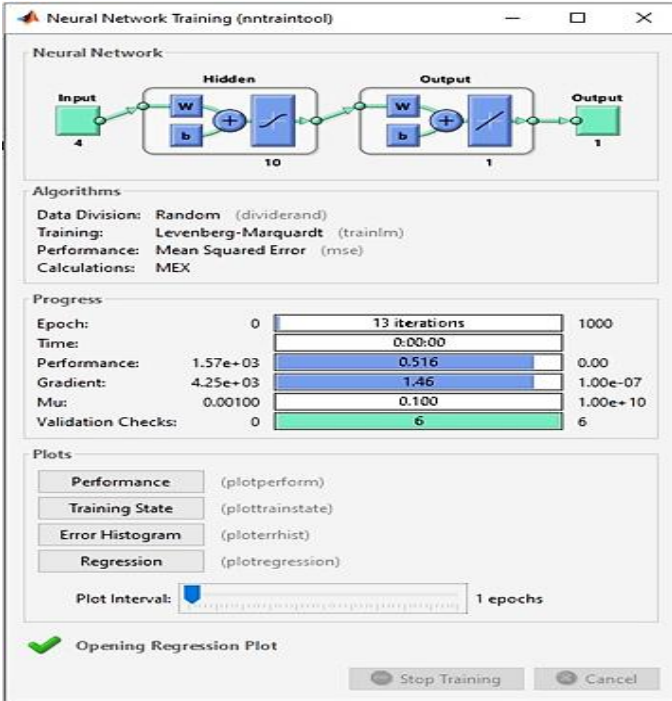

Fig.14: Neural Network Training Interface (Late Fusion)

5.1. Analysis of Neural Network Performance Curves:

Figure (15) shows the performance curves of the late fusion model across the training, validation, and testing phases, with the Mean Squared Error (MSE) tracked over 13 training epochs. The model achieved its best validation performance at epoch 7, recording the lowest MSE value of 8.5057. which indicates that it reached an optimal learning state relatively early. This early drop in error reflects the efficiency of the model's architecture in capturing the fundamental

patterns in the data without requiring extensive training, highlighting the effectiveness of its internal design.

Meanwhile, the training curve continues to decline steadily throughout the learning process, suggesting that the model consistently improves its internal weights and reduces error on the training data, without signs of saturation or stagnation. However, a slight rise in the validation curve after epoch 7 signals the onset of overfitting, where the model begins to adapt too closely to the training data at the expense of its generalization ability. Although this increase is modest, it warrants caution in selecting an appropriate early stopping point to preserve predictive accuracy on unseen data.

The testing curve remains consistently higher than the validation curve, indicating that the model did not generalize perfectly to new data. This gap between validation and testing performance reflects a generalization gap a common phenomenon in models that learn well from original data but struggle with unfamiliar case.

Nevertheless, the overall stability of the curves and the absence of sharp fluctuations suggest that the model maintained balances training behaviour and achieved its lowest error at a suitable point, reinforcing its reliability and highlighting its potential for further improvement through regularization techniques or enhanced training data diversity.
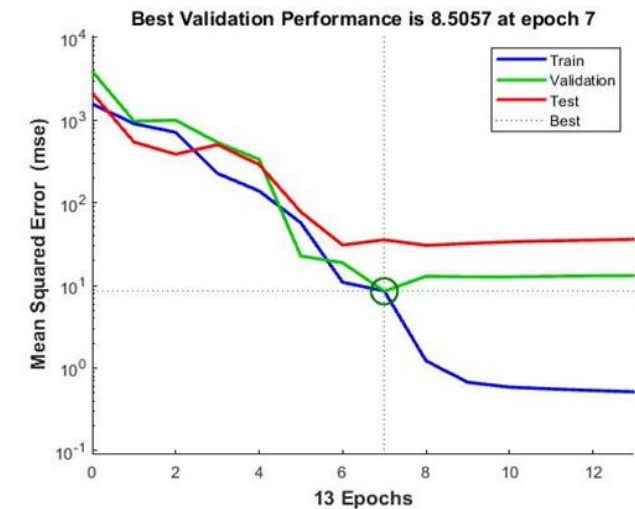

Fig.15: Performance Curves During Neural Network Training (late Fusion)

To address the observed signs of overfitting and improve generalization capability, future studies are encouraged to apply regularization strategies such as Dropout, L2 weight decay, or cross-validation. These techniques can help stabilize learning, reduce variance, and enhance the robustness of multi-output neural architectures under diverse experimental conditions.

5.2. Analysis of Correlation Coefficient (R) Between Predicted and Actual values:

Figure (16) shows four regression plots that illustrate the performance of the late fusion model across the training, validation, testing, and overall phase. These plots depict the relationship between the target values and the model's predicted outputs, each represented by a linear approximation equation and a correlation coefficient (R) that quantifies the strength of the linear fit. This type of analysis serves as a powerful tool for evaluation how accurately the model simulates the true behavior of the data across different stages.

During the training phase, the model achieved an exceptionally high correlation coefficient (R=0.99645), with a regression equation of Output = 0.95×Target+ 7.4 .

This indicates near-perfect alignment between predicted and actual values, as the slope approaches unity and the intercept remains minimal. Such performance reflects the model's strong ability to learn the internal patterns of the data with remarkable precision.

In the validation phase, the model maintained high accuracy, with R=0.99042 and a regression equation of Output = 1 ×Target+ 0.43 . This equation is nearly ideal, suggesting that the model not only learned effectively but also remained stable when exposed to unseen data that was not directly used during training.

In contrast, the testing phase revealed a relative decline in performance.

The correlation coefficient dropped to 0.93224, and the regression

equation shifted to Output = 0.74×Target+ 47 . The reduced slope and increased intercept indicate that the model struggled to generalize perfectly to entirely new data, revealing a generalization gap that may require future improvements in model architecture or training data diversity.

When considering the overall performance across all phases, the model achieved a strong correlation (R=0.99328) with a regression equation of Output = 0.94×Target+ 9.2 .

This demonstrates that the model maintained consistent and reliable behavior in representing the input-output relationship, even when aggregation across varied data conditions.

Based on these findings, it can be concluded that the late fusion model exhibited excellent performance during training and validation, with high precision and stability.

While the testing phase showed room for improvement in generalization, the overall consistency in regression slopes and the high correlation values across phases confirm the model's balanced learning behavior.
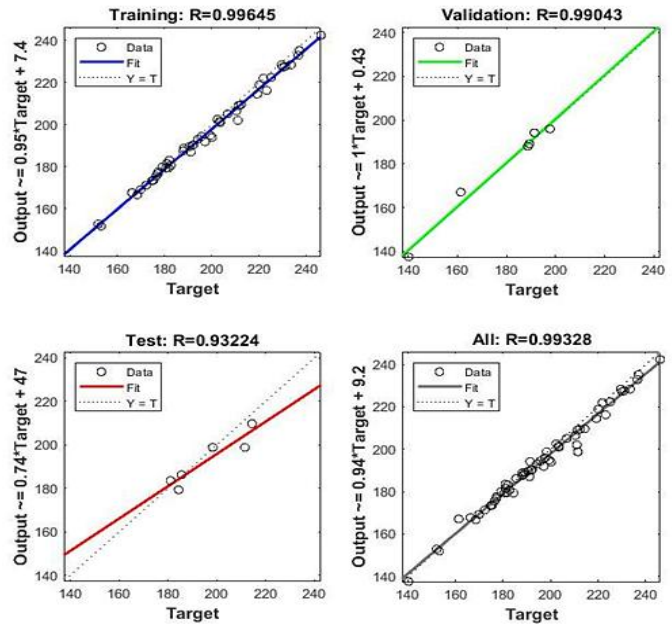

Fig.16: Correlation Coefficients (R) (Late Fusion)
Table (6) provides a summary of the output results.

Table 6: Correlation Coefficients (R) (Late Fusion)

Phase	Correlation Coefficients (R)	Approximation Equation	General Observation
Training	0.99645	Output=0.95×Target+7.4	Very high
Validation	0.99043	Output=1×Target+0.43	High and precise
Test	0.93224	Output=0.74×Target+47	Lower accuracy, needs improvement
All	0.99328	Output=0.94×Target+9.2	Strong overall performance

5.3. Analysis of Training Indicators (Gradient, Mu, Validation Checks) (late Fusion):

Figure (17) shows a comprehensive view of the training dynamics of the late fusion model across 13 epochs, highlighting three key indicators:

Gradient descent behavior, the adaptive learning rate parameter (Mu), and validation checks. Together, these plots offer valuable insight into the model's internal optimization process and its ability to maintain balanced learning.

The first plot illustrates the evolution of the gradient values, which began at a relatively high level and decreased progressively to 1.457 by epoch 13.

This steady decline reflects the model's successful optimization of its internal weights, with no abrupt fluctuations or instability. Such behavior is indicative of a well-structured learning process, where the model continuously minimizes the error function through smooth gradient descent, avoiding erratic updates that could compromise convergence.

The second plot tracks the Mu parameter, which represents the adaptive learning rate within the training algorithm. Mu increased gradually to a value of 0.1 by the final epoch, remaining within the

active adjustment range. This suggests that the model was still refining its parameters but had begun approaching a stable state. The moderate value of Mu indicates that the algorithm was neither stagnant nor overly aggressive, allowing for controlled and meaningful updates without requiring drastic reconfiguration.

The third plot shows the progression of validation checks, which reached the maximum threshold of 6 at epoch 13. This triggered the early stopping mechanism, a safeguard designed to halt training once the model achieves optimal validation performance. The activation of early stopping at this point confirms that the model had reached its peak generalization capacity before entering the overfitting zone, preserving its ability to perform well on unseen data.

Taken together, these indicators demonstrate that the late fusion model exhibited a highly balanced and effective training behavior. The gradual reduction in gradient values, the controlled rise in Mu, and the timely activation of early stopping all point to a model that leaned efficiently, stabilized appropriately, and avoided overfitting. This reinforces the model's robustness and reliability.
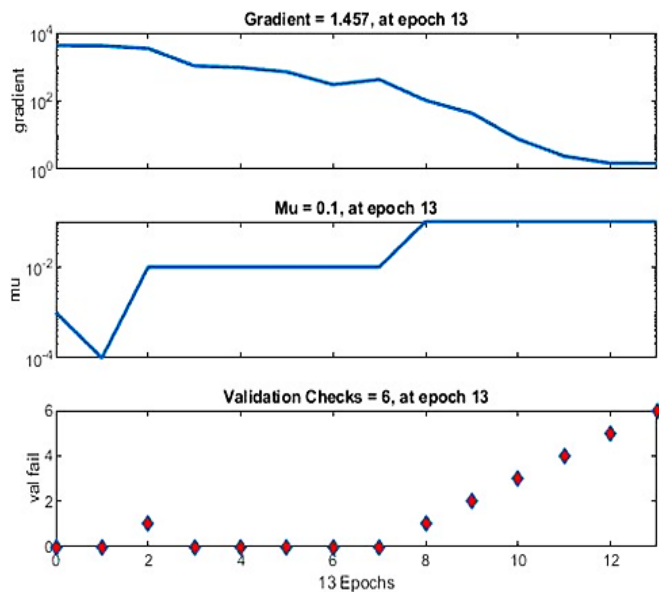

Fig.17: Training Indicators (Gradient, Mu, Validation Checks) (late Fusion)

5.4. Analysis of Error Histogram (Late Fusion):

Figure (18) shows the error distribution plot for the late fusion model, where prediction errors (defined as the difference between target values and model outputs) are grouped into 20 bins across the training, validation, and testing datasets. This type of analysis provides a precise assessment of the model's predictive accuracy and its ability to generalize across different data subsets.

The plot reveals that the majority of errors in the training data are concentrated around the bin centered at 0.24, which lies closest to the zero-error line (indicated by the vertical orange line). This dense clustering near zero reflects the model's high accuracy in prediction, demonstrating its ability to represent the data with minimal deviation. Such behavior indicates that the learning process effectively minimized discrepancies between actual and predicted values.

The overall error range spans from (-5.277 to 12.19), which is relatively moderate and suggests that the model did not produce extreme or erratic predictions. Moreover, the distribution of errors across the training (blue), validation (green), and testing (red) sets appears relatively balanced, indicating that the model performance consistently across different data groups and does not over fit to a specific subset.

This balance in error distribution, combined with the central concentration near zero, serves as strong evidence that the model is not biased or over fitted. Instead, it exhibits stable and well-regulated behavior, reinforcing its reliability for practical applications. Additionally, the absence of skewed distribution or heavy tails suggests that the model is not disproportionately affected by or rare case.
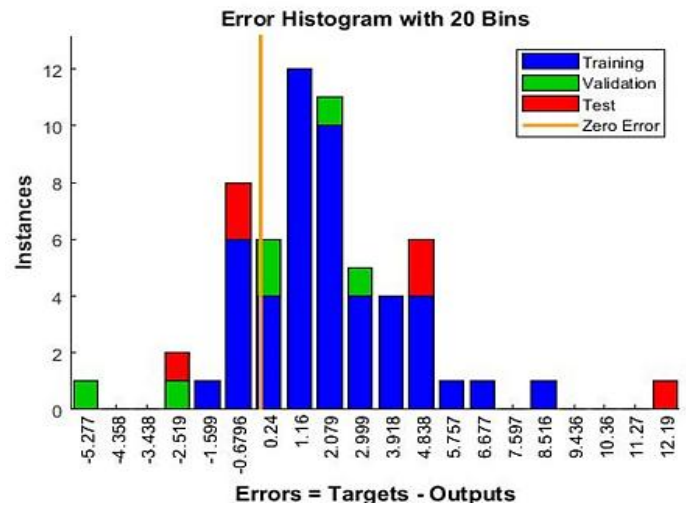

Fig.18: Error Histogram (late Fusion)

5.5. Analysis of Actual vs. Predicted Value Alignment:

Figure (19) shows a precise visual comparison between the actual and predicted values generated by the late fusion model across two distinct outputs: response and response time. The plots reveal a strong visual alignment between the solid black lines (actual values) and the dashed lines (predicted values), demonstrating the model's ability to accurately simulate the true behavior of the data.

In the upper plot, representing the response output, the model closely tracks the variations in the actual data, with the predicted values (blue dashed line) showing near-perfect alignment with the true values (black line). This strong visual match indicates that the model successfully captured the sensory relationship between inputs and outputs without significant deviation, reinforcing its reliability in predicting perceptual behavior.

In the lower plot, which represents response time, the model also exhibits balanced visual performance, The predicted values (red dashed line) closely follow the actual values across different cases. Although time prediction is typically more challenging due to higher variability, the model maintained a commendable level of visual accuracy, suggesting that its learning mechanism effectively captured temporal patterns.

This visual consistency across both outputs serves as strong evidence of the model's generalization capability. The performance was not limited to a specific dataset but extended to unseen case, demonstrating robustness and adaptability. Moreover, the uniformity in prediction quality across different output types reflects internal balance in the model's architecture, further supporting its suitability for multidimensional applications.

Based on these observations, it can be concluded that the late fusion model exhibited visually balanced and effective performance, with clear predictive accuracy and generalization ability.
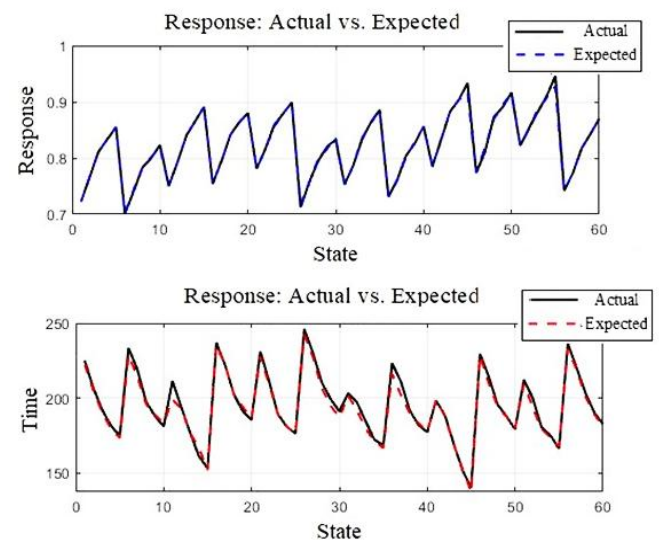

Fig.19: Late Fusion Model Performance Based on Actual vs. Predicted Values

5.6. Scientific Formulation for Publication: Performance Analysis of the Late Fusion Model:

Table (7) shows the performance evaluation of the early fusion model in predicting both response and response time across two phase (training/validation and testing using unseen ratios).

Table 7 :Accuracy Indicators for the Late Fusion Model

Sensor	Dataset Type	Metric Type	MAE	RMSE	R^2	MAPE
Late fusion	Training/ Validation	**(Response)**	0.002	0.004	0.996	%0.19
		Time	2.455	3.356	0.978	%1.23
	Test (Unseen Ratios)	**(Response)**	0.014	0.019	0.841	%1.76
		Time	7.36	10.11	0.674	%3.76

The performance of the late fusion model was evaluated using a set of standard statistical metrics, including (MAE, RMSE, R^2, MAPE). These metrics were applied to both the training/validation dataset (seen data) and the test dataset (unseen data) to assess the model's generalization ability and predictive accuracy across different contexts.

Results indicated that the model achieved high performance on the seen data. For the response variable, MAE was 0.00166 and RMSE was 0.00362, with a very high R^2 value of 0.99603, suggesting that the model explains over 99% of the variance in the data. The MAPE was extremely low (0.19%), reflecting excellent predictive precision. For the time variable, the performance was also satisfactory, with MAE of 2.45502, RMSE of 3.35605, R^2 of 0.97821, and MAPE of 1.23%.

In contrast, the model showed a noticeable decline in performance when applied to unseen data. For the response variable, MAE increased to 0.0142 and RMSE to 0.0194. while R^2 dropped to 0.8417, indicating reduced explanatory power. MAPE also rose to 1.76%. For the time variable, the performance was significantly lower, with MAE of 7.36, RMSE of 10.11, R^2 of 0.6749, and MAPE of 3.76%.

These finding highlight clear challenges in the model's generalization capability, particularly in predicting time under unseen conditions. It is recommended to revisit the data distribution used during training and consider applying regularization techniques such as Dropout or Early Stopping. Additionally, redesigning the model architecture may enhance its generalization ability and improve performance in unseen environments.

Part 2: Comparative Analysis of the Three Neural Models for Predicting Sensors Response and Response Time

Accurate and efficient prediction of response and response time in gas sensor depends on designing neural models that can precisely represent output behavior both numerically and visually. Given these considerations, a structured comparison was carried out to evaluate the performance of the three neural models previously examined-namely, the individual models, early fusion, and late fusion-in order to identify the most reliable and effective architecture.

✓ **Numerical and Training Performance Indicators for the Models (Individual-Early Fusion-Late Fusion):**

The numerical and training performance indicators for the individual, early fusion, and late fusion models are summarised in Table 8.

Table 8: Numerical and Training Performance Indicators

Sensor/Model	Best MSE Achieved	Epoch	Final Gradient	Final Mu	Validation Checks
(SnO_2+WO_3)	0.502	3	$1.22e^{-12}$	$1e^{-11}$	5
(SnO_2+ZnO)	4.999	5	$5.58e^{-13}$	$1e^{-11}$	3
(WO_3+ZnO)	7.285	14	0.008	$1e^{-5}$	6
Early Fusion	17.537	11	0.012	$1e^{-5}$	6
Late Fusion	8.506	7	1.46	0.1	6

✓ **Analysis of Correlation Coefficient (R) and Regression Equations for the Models (Individual-Early Fusion-Late Fusion):**

The correlation coefficients (R) and regression equations for the individual, early fusion, and late fusion models are presented in Table 9.

Table 9: Correlation Coefficient (R) and Regression Equations

Sensor/Model	Prominent Regression Equations	Highest (R) Value
(SnO_2+WO_3)	Output=1×Target+0.01	0.99999
(SnO_2+ZnO)	Output=1×Target+0.011	1
(WO_3+ZnO)	Output=1×Target-2.8e^{-6}	1
Early Fusion	Output=1×Target-0.00036	1
Late Fusion	Output=0.95×Target+7.4	0.99645

✓ **Error Distribution Analysis (Error Histogram) for the Models (Individual-Early Fusion-Late Fusion):**

The error distribution analysis for the individual, early fusion, and late fusion models is presented in Table 10.

Table 10: Error Distribution

Sensor/Model	Error Concentration Location	Overall Distribution Evaluation
(SnO_2+WO_3)	-0.103	Highly accurate and balanced
(SnO_2+ZnO)	-0.048	Slight adjustment needed
(WO_3+ZnO)	+0.3861	Clear deviation
Early Fusion	-0.2483	Accurately distributed
Late Fusion	+0.24	Low and balanced error

✓ **Visual Analysis of Actual vs. Predicted Value Alignment for the Models (Individual-Early Fusion-Late Fusion):**

The visual analysis of actual versus predicted value alignment for the models is summarised in Table 11.

Table 11: Response and Response Time Matching Quality

Sensor/Model	Response Match	Response Time Match	General Observation
(SnO_2+WO_3)	Very High	Excellent	Strong predictive model
(SnO_2+ZnO)	Very Good	Good with minor deviation	Slight fluctuation
(WO_3+ZnO)	Relatively Good	Moderate with clear fluctuation	Needs improvement
Early Fusion	Very High	Good to Very Good	Excellent visual balance, with minor variability
Late Fusion	Very High	Good to Very Good	Strong response match with slight time deviation

❖ **Analysis of prediction Accuracy Discrepancies between Sensitivity and Response Time across all models:**

A comparative evaluation of the three neural architectures (individual models, early fusion, late fusion) reveals a consistent discrepancy in prediction accuracy between the two target variables: sensitivity and response time. This indicates that the models did not predict both variables with equal precision.

In the individual models, the (SnO_2+WO_3) configuration demonstrated superior accuracy in predicting sensitivity (R^2 = 0.91791) compared to response time (R^2 = 0.83882). Conversely, the (WO_3+ZnO) model achieved higher accuracy in predicting response time (R^2 = 0.92471) than sensitivity (R^2 = 0.91974). The (SnO_2+ZnO) model showed the weakest generalization for sensitivity (R^2 = 0.53324), despite maintaining strong performance in response time prediction (R^2 = 0.91068).

The early fusion model exhibited acceptable performance in predicting response time (MAE=2.51012), but its accuracy in sensitivity prediction was notably lower (MAE=0.05338). This suggest that early fusion may introduce representational overlap between tasks, limiting the model's ability to specialize in each output.

In contrast, the late fusion model demonstrated the most balanced performance, achieving exceptionally high accuracy in sensitivity prediction (MAE=20.00166) and relatively strong results for response time (MAE=2.45502). This supports the effectiveness of task separation in enhancing predictive precision across distinct outputs.

Potential causes for this discrepancy include:

• Variable nature: Sensitivity tends to follow more stable patterns based on material properties, while response time is influenced by dynamic factors such as adsorption/desorption rates and internal diffusion, making it more volatile and harder to model.

• Model architecture: Task-specific models (late fusion) allow for deeper specialization, whereas unified models (early fusion) may suffer from representational interference.

• Data distribution: Sensitivity data may exhibit more consistent trends, facilitating learning, while response time data may contain greater variability across samples.

Future studies should consider designing neural architectures that incorporate task (specific optimization) such as modular subnetworks or late fusion strategies to ensure balanced predictive performance across multiple outputs and improve model reliability in real-world sensing applications.

❖ **Statistical Analysis of Accuracy Indicators (MEA, RMSE, R^2, MAPE) for the Models (Individual-Early Fusion-Late Fusion):**

The individual model based on the (SnO_2+WO_3) sensor demonstrated

superior numerical accuracy in sensitivity prediction, achieving the lowest MAE and RMSE values on unseen data, along with a significant improvement in R^2 from 0.67 to 0.92.

Its MAPE value of 1.33% further confirms its reliability in extrapolating to novel doping ratios. In contrast, the (WO₃+ZnO) model showed the highest R^2 (0.92471) and lowest MAPE (1.84%) for response time prediction, indicating strong temporal modelling performance. The (SnO2+ZnO) model, however, exhibited a notable drop in R^2 for sensitivity (from 0.66 to 0.53) and a higher MAPE (2.48%), Suggesting limited generalization despite acceptable time prediction accuracy.

The early fusion model yielded balanced performance in training, with moderate MAE and RMSE values and R^2 scores above 0.66 for both outputs. However, its generalization on unseen data was weaker, particularly for response time, where RMSE rose to 12.05and R^2 dropped to 0.5586. This suggests that early fusion may introduce representational overlap between tasks, limiting specialization and reducing predictive precision.

In contrast, the late fusion model achieved the most consistent and balance performance across both outputs. It recorded the lowest MAE (0.00166) and RMSE (0.00362) for response prediction in training, with an exceptionally high R^2 of 0.99603 and a minimal MAPE of 0.19%. While its performance declined on unseen data, the model maintained acceptable accuracy levels, with R^2 values of 0.8417 for response and 0.6749 for response time, and MAPE values of 1.76% and 3.76%, respectively.

Overall, the statistical indicator confirm that the late fusion model offers the best trade-off between precision and generalization, making it the most suitable architecture for multi-output prediction tasks. The individual (SnO₂+WO₃) model remains optimal for single-output sensitivity prediction, while the early fusion model, despite its training stability, requires further refinement to improve generalization on unseen data.

❖ **Comparison of Proposed Models with Existing Approaches for Predicting Sensitivity and Response Time:**

A review of prior literature reveals that most AI-based studies have focused on gas classification or leak detection tasks using techniques such as early fusion, convolutional neural networks (CNN), recurrent architectures (GRU, Bi-LSTM), or multitask systems. While these approaches achieved high classification accuracy, they did not explicitly address the quantitative prediction of two core sensor performance metrics: sensitivity and response time.

In contrast, the present study introduces a novel contribution by developing neural network models capable of accurately predicting both variable using three distinct architectures (individual, early fusion, late fusion). The results demonstrate that the late fusion model outperforms existing methods in terms of precision (MAE, RMSE, R^2), offering balanced and stable predictions for both outputs. Unlike previous studies, this model separates tasks effectively and provides detailed quantitative evaluation, which enhances its generalization capability and practical relevance.

Therefore, the proposed models go beyond conventional AI applications by offering a robust framework for multi-output prediction in intelligent sensing system, representing a meaningful advancement in both research and industrial contexts.

6.   Limitations of the study:

We acknowledge several limitations in the current study:

First, all experiments were conducted at a fixed gas concentration of 500 ppm, which may not reflect sensor behaviour under varying exposure levels.

Second, only two gas types (acetone, ethanol) were evaluated, limiting the generalizability of the models to broader chemical environments.

Third, the operating temperature were restricted to (250^0C, 350^0C), which may not capture performance dynamics at lower or higher thermal conditions.

Additionally, environmental factors such as humidity, cross-gas interference, and long-term drift were not considered in this phase. These limitations are recognized as areas for future expansion to enhance model robustness and practical applicability.

7.   Conclusion

This study successfully demonstrated the application of artificial intelligence (specifically neural network architectures) in predicting two critical performance metrics of gas sensors: sensitivity and response time. Three sensor configurations (SnO2+WO₃, SnO2+ZnO, WO₃+ZnO) were fabricated and evaluated under controlled conditions, and three modelling strategies were implemented: individual models, early fusion, and late fusion. The results revealed that the SnO2+WO₃ sensor model achieved the highest accuracy in sensitivity prediction, while the WO₃+ZnO model excelled in response time estimation. The late fusion architecture emerged as the most balanced and reliable approach, offering superior generalisation and predictive stability across both outputs. In contrast, the early fusion model showed good training performance but limited generalisation, particularly for time prediction. Statistical indicators (MAE, RMSE, R^2, MAPE), regression analysis, error histograms, and visual alignment plots collectively confirmed the robustness of the late fusion model and its suitability for multi-output prediction tasks.

Moreover, the neural networks uncovered nonlinear relationships between doping ratios, temperature, and gas type, providing deeper insights into sensor behaviour that traditional analysis could not reveal.

8.   Future Work:

Building on the current findings, future work will focus on expanding the scope and applicability of the proposed models.

First, we plan to explore more advanced neural architectures such as convolutional neural networks (CNNs). recurrent neural networks (RNNs), and GRU-based models to capture temporal and spatial dependencies in sensor data [7].

Second, the study will be extended to include a broader range of gases, including ammonia, nitrogen dioxide, and volatile organic compounds (VOCs), to improve generalizability across industrial and environmental contexts.

Third, real-world datasets will be incorporated to evaluate model robustness under uncontrolled conditions, including variable humidity, temperature fluctuations, and cross-gas interference [16].

Finally, application-specific optimization will be pursued, such as tailoring models for wearable sensors, smart city monitoring, and food safety applications, where rapid and reliable gas detection is critical [17]. These directions aim to enhance both the scientific depth and practical relevance of intelligent gas sensing system.

9.   References

[1]   Narkhede, Parag., Walambe, Rahee., Mandaokar, Shruti., Chandel, Pulkit., Kotecha, Ketan., & Ghinea, George. (2021). Gas Detection and Identification Using Multimodal Artificial Intelligence Based Sensor Fusion. Applied System Innovation, 4(1), 3. https://doi.org/10.3390/asi4010003

[2]   Walambe, Rahee., Narkhede, Parag., Mandaokar, Shruti., et al. (2023). Gas Detection and Identification Using Multimodal AI-Based Sensor Fusion. Data Science Week Poster Presentation.

[3]   Zhang, Yue., Li, Hong., Chen, Xiaohui., & Wang, Jian. (2024). Gas Detection and Classification Using Multimodal Data Based on Deep Learning. Sensors, 24(18), 5904. https://doi.org/10.3390/s24185904

[4]   Chakraborty, Sourav., Mittermaier, Stefan., & Carbonelli, Claudio. (2022). Understanding the Behavior of Gas Sensors Using Explainable AI. Engineering Proceedings, 27(1), 61. https://doi.org/10.3390/ecsa-9-13350

[5]   Zhuang, Yuxuan., Yin, Dong., Wu, Lei., Niu, Guanghui., & Wang, Feng. (2024). A deep learning approach for gas sensor data regression: Incorporating surface state model and GRU-based model. APL Machine Learning, 2(1), 016104. https://doi.org/10.1063/5.0160983

[6]   Attallah, Omar., & Morsi, Islam. (2024). Multitask Deep Learning-Based Pipeline for Gas Leakage Detection via E-Nose and Thermal Imaging Multimodal Fusion. Chemosensors.

[7]   Liu, Ming., & Zhao, Qiang. (2024). 2024 Breakthroughs in

Smart Gas Sensor Technology: A Review. Gas Detection Journal. Retrieved from https://gasdetection.com/articles/2024-breakthroughs-in-smart-gas-sensor-technology-a-review/.

[8] Ahmed, Shahbaz., Kumar, Rajesh., & Lee, Donghyun. (2024). Assessing Gas Leakage Detection Performance Using Machine Learning with Different Modalities. Journal of Intelligent & Fuzzy Systems, 45(2), 123–135. https://link.springer.com/article/10.1007/s42341-024-00545-0

[9] Angstenberger, Stefan., Sterl, Florian., Theelke, Kilian., Giessen, Harald., & Schwarz, Ulrich. (2025). Real-time detection of low gas concentrations using coherently controlled quartz-enhanced photoacoustic spectroscopy. Optica, 12(1), 45–52. https://doi.org/10.1364/OPTICA.544448.

[10] Al-Hadeethi, Yousef., Alghamdi, Abdullah., Alshahrani, Hassan., Alqahtani, Mohammed. and Alshahrani, Fahad., 2023. Fabrication of ZnO/WO$_3$ nanocomposite thin films using sacrificial template method for enhanced VOC gas sensing. Nanomaterials, 13(4), p.733. Available at: https://doi.org/10.3390/nano13040733

[11] Korotcenkov, Ghenadii., Brinzari, Valeriu., Cho, Byung-Keun., Gulina, Larisa. and Han, Seung-Hwan., 2014. The influence of heterojunctions on the gas sensing properties of SnO$_2$–WO$_3$ composite thin films. Sensors, 14(11), pp.20480–20512. Available at: https://doi.org/10.3390/s141120480

[12] Zhang, Jian., Liu, Xiaobo., Neri, Giovanni. and Pinna, Nicola., 2015. Nanostructured materials for room-temperature gas sensors. Nanoscale, 7(6), pp.2028–2043. Available at: https://doi.org/10.1039/C5NR02571K.

[13] Al-Hadeethi, Yousef., Alshahrani, Hassan. and Alqahtani, Mohammed., 2024. Enhanced NO$_2$ gas sensing performance of SnO$_2$-WO$_3$ nanocomposite thin films prepared by chemical bath deposition. Journal of Electronic Materials. Available at: https://doi.org/10.1007/s11664-024-11381-6

[14] Hosseini-Golgoo, Seyed Mohammad., Yousefi, Mohammad Hossein., & Ghasemi, Saeid., 2013. Application of neural networks for gas identification using temperature-modulated sensors. Sensor Letters, 11(3), pp.556–561.

[15] Zhou, Yue. and Liu, Xiaobo., 2021. Gas recognition based on time-series neural networks and sensor arrays. Sensors, 21(14), p.4826. https://doi.org/10.3390/s21144826

[16] Dehnaw, Ali. et al., 2024. Deep Neural Network Optimization for Efficient Gas Detection Systems in Edge Intelligence Environments. Processes, 12(12), p.2638. https://doi.org/10.3390/pr12122638

[17] Zong, Boyang. et al., 2025. Smart Gas Sensors: Recent Developments and Future Prospective. Nano-Micro Letters, 17(54). https://doi.org/10.1007/s40820-024-01543-w