Indonesian Journal of Electrical Engineering and Computer Science
Vol. 40, No. 3, December 2025, pp. 1325~1336
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v40.i3.pp1325-1336

Meta-model integration with attention mechanisms for advanced decision-level fusion in machine learning

Shobha, Nalini Narasimhaiah
Department of Computer Science and Engineering, Nitte Meenakshi Institute of Technology,
Affiliated to Visvesvaraya Technological University, Bengaluru, India

Article Info

Article history:

Received Dec 15, 2024
Revised Jul 3, 2025
Accepted Nov 16, 2025

Keywords:

Attention mechanism
Data fusion
Decision-level fusion
Ensemble learning
Machine learning
Meta-models
Prediction accuracy

ABSTRACT

This work proposes an advanced meta-model approach that incorporates forecasts from multiple machine learning models to improve classification accuracy in complex tasks. The approach employs decision-level data fusion, where predictions from random forest (RF), XGBoost, neural networks (NN), and support vector machine (SVM) are combined within a meta-model framework. The meta-model incorporates an attention mechanism and a gated model selection process to dynamically emphasize the most relevant model outputs based on input features. The results demonstrate superior accuracy in predicting explicit content compared to traditional fusion methods. This research highlights the potential of attention-enhanced meta-models in improving interpretability and accuracy across various domains. The integration of meta-models with attention mechanisms has the potential to significantly enhance decision-level fusion in machine learning applications. This study investigates the development of an advanced fusion framework leveraging attention mechanisms to improve decision-making accuracy in multi-source data environments. The proposed method is evaluated across multiple datasets, demonstrating its efficacy in increasing predictive performance and robustness.

Corresponding Author:

Shobha
Department of Computer Science and Engineering
Nitte Meenakshi Institute of Technology
Bengaluru, Affiliated to Visvesvaraya Technological University, India
Email: shobha.p@nmit.ac.in

1. INTRODUCTION

In the period of machine learning and artificial intelligence, data fusion has emerged as a essential methodology for enhancing predictive accuracy, particularly when integrating multiple data sources. Decision-level data fusion, a widely adopted approach, aggregates outputs from various models to enhance classification performance. While traditional methods such as majority voting and weighted voting have been effective, they often fail to optimize model contributions dynamically. The proposed research introduces a novel meta-model with an attention mechanism to address this limitation, safeguarding adaptive weighting of model outputs for improved predictive accuracy. Decision-level data fusion involves the combination of decisions or predictions from multiple models or sources to reach a conclusion. Unlike data-level and feature-level fusion, which operate on raw data and features, decision-level fusion works with the outputs of different models or decision-making processes. The main benefit of decision-level fusion lies in its capacity to integrate different viewpoints from multiple models, each potentially trained on distinct aspects or subsets of the data. This type of fusion is especially valuable in situations where data sources are various or when models are established using various approaches, such as statistical techniques, machine learning (ML) algorithms, or expert systems. There are numerous methodologies for implementing decision-level data fusion, each offering distinct advantages and suited to specific applications. The most common methods include majority voting, weighted voting, and the use of meta-models. In the majority voting approach, each model casts a "vote" for a particular decision or prediction. The final decision is made based on most votes. This method is simple and effective, particularly when the models are equivalent in terms of their predictive power. Majority voting is often used in ensemble learning techniques, where multiple models are fused to improve accuracy.

Weighted voting extends the majority voting method by assigning a weight to each model's vote, reflecting its estimated accuracy or/and reliability. This approach ensures that models with higher reliability have a more significant impact on the final decision. Models that have historically performed better on a given task are given more influence in the final decision. This method permits more nuanced decision-making, particularly in cases where certain models are known to perform better on specific subsets of data. A more sophisticated approach to decision-level fusion involves the use of meta-models, also known as stacked generalization or ensemble learning. In this method, the outputs of the base models are used as inputs to a higher-level model, which is then trained to make the final decision. The meta-model can learn to weigh the predictions of the base models based on their performance, potentially using techniques like attention mechanisms or gating mechanisms to dynamically adjust the influence of each base model. This method can substantially improve predictive performance, particularly in complex tasks where different models capture various features of the data.

Decision-level data fusion is broadly used in several fields, including healthcare, radar data set, finance, remote sensing, and autonomous systems. In the healthcare sector, decision-level fusion is used to combine predictions from various diagnostic models, such as those used in medical imaging, patient monitoring, and genomic analysis. By fusing the outputs of these models, health care organizations can make more precise diagnoses and treatment decisions, reducing the likelihood of false positives or negatives. In cancer detection, models trained on imaging data, patient history, and genomic data can be fused to provide a more comprehensive assessment of a patient's condition. In finance, decision-level fusion is employed in credit scoring, fraud detection, and algorithmic trading. Different models analyze various aspects of financial data, such as transaction history, market trends, and social media sentiment. By fusing these decisions, financial institutions can make more informed and robust decisions, such as approving a loan or flagging a potentially fraudulent transaction. In remote sensing, decision-level fusion is used to integrate data from multiple sensors, such as satellites, drones, and ground-based sensors. This approach is particularly beneficial in applications such as environmental monitoring, where various sensors contribute complementary data, enhancing the overall accuracy of the system. Satellite imagery provide a broad overview of an area, while drone footage offers high-resolution details. Combining these data sources at the decision level allows for more accurate and timely assessments of environmental conditions. Autonomous systems, such as self-driving cars and drones, rely heavily on decision-level fusion to integrate inputs from various sensors, such as cameras, LiDAR, and radar. By fusing these decisions, the system can make more reliable decisions about navigation, obstacle avoidance, and path planning. This is crucial in ensuring the safety and efficacy of autonomous vehicles in dynamic and unpredictable environments.

Though decision-level data fusion provides significant advantages, it also causes challenges that need to be overcome to achieve the best possible performance. A key challenge in decision-level fusion is ensuring that the base models exhibit sufficient diversity. If the models are too similar, they tend to make the same errors, which diminishes the effectiveness of the fusion process. Achieving model diversity requires careful selection of models, training data, and feature sets to ensure that each model contributes unique insights to the final decision. Properly weighing the contributions of different models is another challenge. Assigning appropriate weights to each model's predictions requires a thorough understanding of their strengths and weaknesses. The weights must be calibrated over time as the models are exposed to new data. This dynamic weighting process can be complex, particularly in real-time applications where decisions must be made quickly. The meta-model in decision-level fusion must be carefully designed to avoid overfitting the training data. Overfitting occurs when the meta-model becomes too closely tailored to the specific examples it has seen during training, leading to deficient performance on new, unseen data. Ensuring good generalization requires techniques such as cross-validation, regularization, and the use of large and diverse training datasets. Decision-level fusion, particularly when using meta-models, can be computationally intensive. Training multiple base models along with a meta-model demands substantial computational resources, particularly when handling large datasets or complex models such as deep neural networks (DNN). This computational complexity can be a limiting factor in applications where real-time decision-making is required.

As ML continues to evolve, various promising directions for future research and development in decision-level data fusion are evolving. One area of relevance is the development of adaptive fusion techniques that dynamically adjust the fusion strategy based on the characteristics of the data or the performance of the models. An adaptive system increase the weight of a particular model's predictions when it is known to perform well on certain types of data or in specific conditions. This approach could improve the robustness and flexibility of decision-level fusion systems, particularly in dynamic environments. Another important direction for future research is improving the explainability and interpretability of decision-level fusion systems. As these systems become more complex, it becomes increasingly challenging to understand how decisions are being made and why certain models are being weighted more heavily than others. Creating methods that offer insights into the decision-making process, such as visualizations or explanations of the model's reasoning, is essential for building trust and acceptance in critical applications like healthcare and finance. Future exploration may also include the integration of decision-level fusion with data-level and feature-level fusion. By combining multiple levels of fusion, it is possible to achieve even greater predictive accuracy and robustness. A system could first perform data-level fusion to preprocess and clean the data, then apply feature-level fusion to create a rich set of features, and finally use decision-level fusion to combine the outputs of multiple models. This multi-level fusion approach could provide a more holistic and powerful solution to complex problems. As decision-level fusion techniques continue to mature, there is significant potential for their application in new domains. For example, the growing field of internet of things (IoT) could benefit from decision-level fusion to integrate data from diverse sensors and devices, enabling more intelligent and autonomous systems. Similarly, decision-level fusion could play a key role in areas like cybersecurity, where combining the outputs of different detection models could improve the identification and mitigation of threats.

Decision-level data fusion represents a powerful and versatile approach to enhancing predictive accuracy and robustness in ML systems. By combining the outputs of multiple models, each offering a unique perspective on the data, decision-level fusion can produce more reliable and accurate decisions than any individual model could achieve on its own. While this method presents several challenges, including the need for model diversity, proper weighting, and computational efficiency, ongoing research and development are poised to address these issues and unlock new possibilities for decision-level fusion in a wide range of applications. As the field progresses, decision-level fusion is expected to become increasingly vital in empowering intelligent systems to adapt to intricate and ever-changing environments, making it a crucial area of focus for the future of AI and ML. A contrastive learning framework enhances few-shot text classification with improved performance. A contrastive learning framework with supervised and unsupervised regularization for few-shot text classification, coupled with self-paced episode sampling, significantly improves performance and training efficiency over existing methods [1]. Standard deep learning models, when properly tuned, can achieve comparable or superior performance to meta-learning models for cold-start problems in recommender systems [2]. RESUS [3] uses a shared predictor for global preferences and efficient algorithms for individual preferences, outperforming existing methods in accuracy and efficiency on cold users. The techniques like S-CNF generate diverse, human-like annotations efficiently and excels in emotion recognition and toxic speech detection [4], [5]. The ML techniques can be applied for EEG features and demographic data to diagnose schizophrenia (SCZ). The model [6], trained on a small Chinese dataset and enhanced with a large American dataset, shows robust cross-site performance and identifies EEG theta and alpha band power as key biomarkers. The approach demonstrates the feasibility of using EEG for SCZ diagnosis across diverse clinical settings. Blockchain, AI, and digital twin technologies are incorporated to enrich the metaverse. By examining advancements in digital currencies, AI applications, and digital twins, the chapter highlights the need for interdisciplinary collaboration to create an open, fair, and innovative metaverse [7]. NeuEvo, a biologically inspired approach, integrates and feedback connections with excitatory and inhibitory neurons in spiking NNs. By using spike-timing-dependent plasticity and global error signals, it evolves neural circuits, boosting implementation in image classification and reinforcement learning tasks. NeuEvo achieved state-of-the-art results on several datasets and matches artificial NNs, paving the way for more complex network structures [8].

Standard deep learning methods, relying on fixed phases of weight updates and evaluations, struggle with continual learning, leading to a loss of plasticity over time. In contrast, algorithms like continual backpropagation, which randomly initialize less-used units, maintain plasticity indefinitely. The findings suggest that sustaining deep learning requires integrating random components alongside gradient descent to preserve variability and plasticity [9]. This study proposed by Zhang and Huang [10] evaluates the impact of large language model (LLM)-based Chatbots on foreign language students' vocabulary acquisition. Over eight weeks, 52 students used either a Chatbot or traditional methods to learn target words. Assessments revealed that Chatbots significantly improved both receptive and productive vocabulary knowledge, enhanced long-term retention, and facilitated incidental learning. The findings underscore the effectiveness of LLM-based tools in language education, highlighting their potential for vocabulary development and encouraging their integration into teaching strategies. The integration of AI into pharma is transforming drug discovery, development, and digital health. This shift includes leveraging AI for drug target discovery, novel compound design, efficient clinical trials, and optimizing manufacturing and marketing. Pharma is also exploring AI in digital health for diagnostics, personalized treatments, and patient monitoring. Key areas of focus include digital therapeutics, telehealth, and secure health data management. The paper discusses technology and market drivers, as well as trends shaping AI's role in pharma's digital transformation [11]. The rise of Internet-of-Things and smartphones has spotlighted speech emotion recognition (SER) technologies, yet their adoption faces conceptual, technical, and societal hurdles. Challenges include the complexity of human emotion, obtaining reliable data, and diverse contextual factors. Societal concerns about privacy, fairness, and explainability also impact acceptance. This article reviews the history and current advancements in SER and explores algorithmic approaches to make these technologies more accessible and responsible [12], [13]. WhisperX addresses the limitations of large-scale, weakly-supervised speech recognition models like Whisper by providing time-accurate transcription with word-level timestamps. Utilizing voice activity detection and forced phoneme alignment, WhisperX enhances long-form transcription and word segmentation, offering state-of-the-art performance. Its valence-arousal-dominance (VAD) cut and merge strategy pre-segments audio, improving transcription quality and achieving a twelve-fold speedup through batched inference [14]. The educational and motivational potential of mobile assisted language learning (MALL) applications explored [15]. It begins with definitions and examples of MALL methodologies, highlighting their advantages and issues identified in existing literature. The thesis then examines popular MALL applications like Duolingo, Babbel, and Memrise, analyzing their features for language learning. Research involving 20 young adults who used a language learning app for three weeks revealed that while the apps are interactive, easy to use, and motivating, they have content deficiencies and repetitive basic activities. Based on these findings and previous research, the thesis proposes methodological improvements for these applications. A technique combining feature engineering, ensemble learning, DNNs, and transfer learning to enhance fusion score levels can be used to increase accuracy, precision, and efficiency while reducing reasoning time and resource consumption across diverse datasets [16]. the INAFEN framework [17] used for logistic regression, that automates feature engineering to transform nonlinear relationships into linear ones and incorporates feature cross and knowledge distillation. The framework achieves high predictive performance and superior interpretability compared to black-box models.

Feature engineering is also used by Zhao B to predict lithium-ion battery life by preprocessing data, engineering 79 features, and selecting 16 key features. It uses a sparse autoencoder and Transformer model to achieve high accuracy, with errors as low as 2.6% for early predictions [18]. Similarly, methods to optimize feature selection and application of an ensemble of diverse models, including support vector machine (SVM), K-nearest neighbors, and XGBoost. The EHMFFL algorithm, tested on Cleveland and Statlog datasets, achieves high accuracy rates of 91.8% and 88.9%, respectively, demonstrating its effectiveness in improving diagnostic accuracy and supporting clinical decision-making. A predictor that integrates human decisions into the final output, showing both theoretical and empirical advantages over traditional methods was discussed by Charusaie *et al*. [19]. For the emotional mimicry intensity (EMI) challenge in Affective Behaviour Analysis in-the-wild, a Wav2Vec 2.0 [20] pre-trained on podcasts to capture diverse audio features. The approach employs a fusion technique that integrates individual features with a global mean vector and incorporates a pre-trained VAD model to enhance emotion intensity prediction across multiple emotional dimensions. SER was introduced by Upadhyay *et al*. [21] incorporating rater subjectivity through perception-coherent clusters (PCC), expanding the label space for better learning. Evaluated on the IEMOCAP and MSP-Podcast corpora, the rater perception coherency (RPC)-based SER model outperformed consensus-based models, achieving 3.39% UAR improvement on IEMOCAP and 2.03% on MSP-Podcast, with detailed insights into the role of perception consistency in SER learning. Accurate confidence estimation enables conversational search systems to tailor responses based on user expertise [22]. A meta-model combining transfer learning, adversarial autoencoder, and variational mode decomposition for structural health monitoring using optimized KNN classifiers discussed [23], [24]. A hybrid fusion model combining early and late data fusion techniques to improve fake news detection by integrating text and images effectively [25]. CNNs and vision transformers combined to detect lung diseases from chest X-rays. CatBoost and bagging models achieved high accuracy, significantly improving classification performance for COVID-19 and pneumonia [26], [27]. AI, IoT, and CPS integration, with 6G technology and data fusion, is transforming Industry 4.0.

Data fusion is crucial for enhancing predictive accuracy, particularly in multi-source ML applications. Decision-level fusion integrates predictions from multiple models, leveraging diverse perspectives for improved classification [28]. Existing methods such as majority voting and weighted voting often fail to optimize model contributions dynamically. This research introduces a meta-model with an attention mechanism to address this challenge. The novelty lies in the adaptive weighting of model outputs, ensuring optimal prediction refinement. This study contributes by providing an interpretable and efficient meta-learning framework. The meta-models, and innovative techniques enhance few-shot text classification, emotion recognition, and AI applications. his study aims to develop a meta-model framework with an integrated attention mechanism that dynamically adjusts model contributions based on their relevance. By doing so, the study seeks to enhance predictive accuracy, improve robustness against noisy or incomplete data, and provide a scalable solution adaptable to various domains, including healthcare, finance, and autonomous systems. Existing studies have explored diverse data fusion techniques. Majority voting methods provide a simple and interpretable approach but lack flexibility in adjusting model contributions dynamically. Weighted voting improves upon this by assigning model-specific weights based on prior accuracy but does not account for real-time feature importance variations. Recent advancements in ensemble learning and meta-learning have introduced stacked generalization and gated mechanisms, yet they often overlook interpretability and adaptability in high-dimensional data environments. Our study bridges this gap by integrating an attention-based meta-model that refines decision fusion dynamically.

The key contributions of this study are as follows,

a) Development of a novel meta-model framework that integrates attention mechanisms for dynamic model selection and weighting.
b) Implementation of a gated mechanism that adjusts model contributions based on feature relevance, ensuring robustness against biases.
c) Comprehensive evaluation across multiple datasets, demonstrating superior classification performance compared to conventional decision-fusion methods.
d) Analysis of computational efficiency, highlighting scalability and adaptability across domains such as healthcare, finance, and autonomous systems.

The remainder of this paper is structured as follows: Section 2 describes the methodology, detailing data preprocessing, model training, and fusion mechanisms. Section 3 presents the experimental setup and evaluation metrics. Section 4 discusses findings, implications, and potential applications. Section 5 concludes with future research directions.

2. METHOD

In the rapidly evolving field of ML, combining multiple models to enhance predictive performance is a novel approach. Figure 1 depicrss a complex ML pipeline that leverages a meta-model to integrate predictions from various base models. The key components of this approach are random forest (RF), SVM, XGBoost, and NNs. Each of these base models brings unique strengths, which are harnessed by the meta-model to improve prediction accuracy, particularly in complex tasks like classifying,

2.1. Feature engineering and model training

The data set used in this study includes podcast metadata collected between 2007 and 2016, consisting of episode length, category, author, language, and show category. Preprocessing steps involve normalization and encoding of categorical features. Base models like RF, SVM, XGBoost, and NN are selected for their complementary strengths. The RF excels in robustness and handling high-dimensional data, SVM is effective in separating complex data distributions, XGBoost offers efficient handling of structured data, and NNs capture intricate non-linear patterns. These models were trained independently using Python’s scikit-learn and TensorFlow libraries on a system with an NVIDIA GPU, with hyperparameters tuned via grid search. The outputs of the base models are integrated using a meta-model with an attention mechanism, which assigns dynamic weights to the model outputs based on their relevance to the final classification. This attention mechanism is trained using backpropagation with a binary cross-entropy loss function, optimized via Adam optimizer. Dropout layers and cross-validation are employed to prevent overfitting. Additionally, a gated mechanism adjusts the contribution of each base model based on data characteristics, enhancing robustness and reducing biases. Evaluation metrics include accuracy, precision, recall, and F1-score, computed via a 5-fold cross-validation technique to ensure reliability. Computational complexity is analyzed by comparing training and inference times across models, highlighting trade-offs between accuracy and efficiency. Exclusion criteria are applied to models with low performance during preliminary tests to ensure only reliable predictions are integrated by the meta-model. The enhanced Method section ensures reproducibility by providing a detailed description of the datasets, model training procedures, meta-model integration, attention mechanism training, evaluation metrics, and computational complexity analysis. The dataset undergoes preprocessing and feature engineering, including episode length, category, and author information. The base models- RF, SVM, XGBoost, and NNs-are trained independently to capture different aspects of the data. The process begins with feature engineering, transformation, and combination. Features such as episode length, category, author, language, and show category are carefully selected and transformed into a format fit for model consumption. This step is vital since the quality of input features directly affects the performance of ML models. Once the features are prepared, they are fed into the following base models - RF, SVM, XGBoost, and NNs. These models are chosen for their complementary strengths. RF is known for its robustness and ability to handle many features.SVM excels in high-dimensional spaces and is effective in cases where the number of dimensions exceeds the number of samples. XGBoost is a powerful boosting algorithm that is highly effective in handling structured/tabular data. NNs are versatile and capable of capturing non-linear relationships in data, making them ideal for complex pattern recognition tasks. Each model makes predictions independently, capturing different aspects of the data.

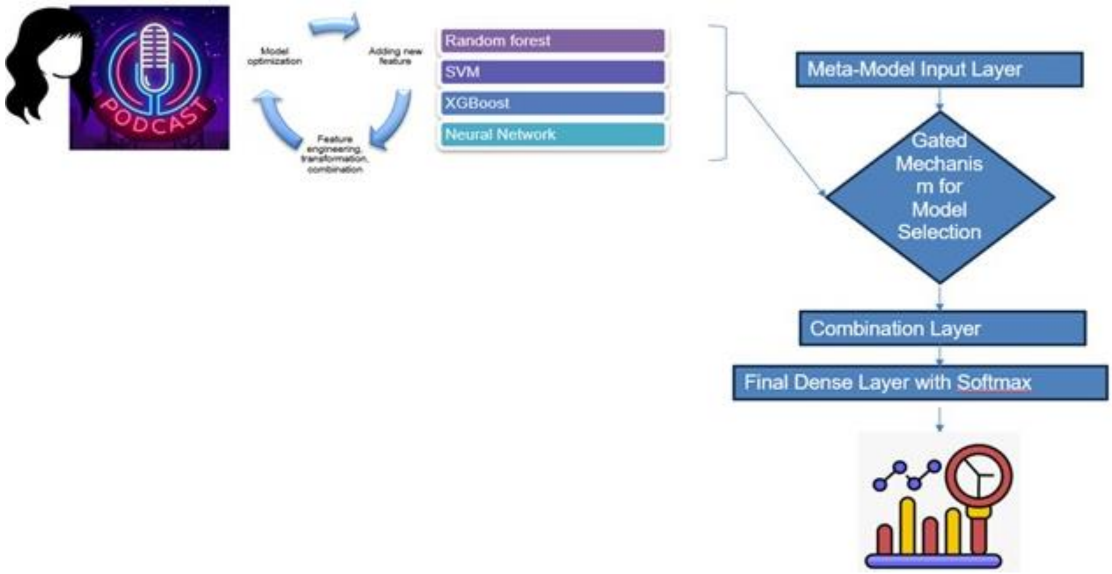


Figure 1. Meta-model approach with attention

2.2. Meta model with attention mechanism

A meta-model aggregates base model predictions using an attention mechanism that assigns dynamic weights based on relevance to the final classification. This increases predictive accuracy while improving interpretability. The predictions from these base models are then combined in a meta-model framework. The meta-model acts as a second-level learner that integrates the outputs of the base models to produce a final prediction. This approach is particularly effective because it allows the meta-model to focus on the strengths of each base model while mitigating their weaknesses. The meta-model in this setup includes an attention mechanism, this is a novel and powerful tool for model selection and combination. The attention mechanism operates by assigning different weights to the outputs of the base models based on their relevance to the final prediction. Essentially, the meta-model "attends" to the most important predictions, enabling a more refined and accurate final output.

2.3. Gated mechanism for model selection

The gated mechanism improves model contributions by dynamically adjusting weights based on data attributes. This mitigates model biases and ensures robustness. In a scenario where the RF model performs well on categorical data but struggles with numerical data, the gated mechanism reduces the weight of the RF's predictions when numerical features are more prominent. Conversely, it increases the weight of the NN's predictions if it excels in capturing the underlying patterns in the numerical data.

The entire process is iteratively optimized through backpropagation, where the model parameters are adjusted to reduce a loss function, and binary cross-entropy in classification tasks are used. The result is a highly accurate and robust predictive model that leverages the strengths of multiple ML algorithms through the power of meta-learning and attention mechanisms. The approach depicted in the image exemplifies the cutting-edge techniques in ensemble learning, particularly the use of meta-models with attention mechanisms. By integrating multiple base models and employing a dynamic, gated selection process, this approach significantly enhances predictive performance, making it a powerful tool in complex ML tasks.

This framework determines improved accuracy in predicting explicit content in podcast episodes, as validated on a test dataset. The integration of attention mechanisms and a gated model selection strategy provides a more flexible and interpretable approach to meta-learning, making it suitable for complex classification tasks in various domains. The proposed method sets a foundation for future research in enhancing meta-learning models through adaptive and focused learning strategies. The proposed process where the base models' predictions are combined and processed by a NN with an attention mechanism, ultimately resulting in a final prediction on whether an episode is explicit or not. The final precision of the meta-model is computed by comparing these predictions to the actual outcomes. Let x be the feature matrix and y be the target variable. x_{train} , x_{test} , y_{train} , y_{test} are the training and testing splits of x and y. $RF(x)$ represents the probability of the RF classifier for input x . $SVM(x)$ represents the probability of output of the SVM. M represents the meta-model, which is a NN with attention mechanisms.

The method of the training the base models if represented by (1) and (2),

$$RF_{train} = RF_{trainx} \tag{1}$$

And,

$$SVM_{train} = SVM_{trainx} \tag{2}$$

The base model predictions are represented by (3), (4), (5) and (6).

$$P_{RF,train} = RF\ (x_{train}) \tag{3}$$

$$P_{SVM,train} = SVM\ (x_{train}) \tag{4}$$

$$P_{RF,test} = RF\ (x_{test}) \tag{5}$$

$$P_{SVM,test} = SVM\ (x_{test}) \tag{6}$$

From the Meta-Model inputs are represented by (7) and (8),

$$x_{meta,train} = [P_{RF,train}, P_{SVM,train}] \tag{7}$$

$$x_{meta,test} = [P_{RF,test}, P_{SVM,test}] \tag{8}$$

In Meta-Model with attention Mechanism,
The input to the meta – model is given (9)

$$z = x_{meta,} \tag{9}$$

The dense layer with ReLU activation,

$$d = ReLU(W_{d^2} + b_d\ ) \tag{10}$$

Apply dropout using (10),

$$d_{drop} = Dropout(d)$$

Reshape for attention is given by $d_{reshape} = Reshape(d_{drop})$.

Attention mechanism,
Let $A = Attention\ (d_{reshape}\ , d_{reshape})$, Global average pooling is given by $f =$ GlobalAveragePooling1D(A).
Final dense layers: $h = ResLU(w_h\, f + b_h)$ Output layer: $\hat{y} = Sigmoid\ (w_o\, h + b_o\ )$. Final Prediction and Accuracy is reprsented by (11).

$$\widehat{y_{meta}} = M(x_{meta,test}) \tag{11}$$

$$\text{Meta model Accuracy} = \frac{1}{N}\sum_{i=1}^{N} 1(y_{meta,i} = y_{test,i}\ ) \tag{12}$$

Where,
$W_d\ , W_h\ , W_o\ and\ b_d\ , b_h, b_o$ are the weights and biases of the respective layers.
1(·) is the indicator function, equal to 1 if the condition inside is true, otherwise 0.

2.4. Learning attention weights

The proposed framework's computational efficiency is evaluated by analyzing training and inference costs across the base models and meta-model. The Base models are trained, and their predictions combined. A meta-model with attention mechanism processes these predictions to make final predictions, with accuracy evaluated against actual outcomes. The attention mechanism assigns weights to the outputs of base models to emphasize predictions most relevant to the final output. This weight assignment is not arbitrary; it is learned through an optimization process as part of the meta-model training. Specifically, attention weights $\alpha_i$for each base model i are calculated dynamically based on the feature set X of the input data. Feature Representation Transformation: The meta-model first maps the input features X into a higher-dimensional representation using a dense layer with trainable weights WW and biases b.

$$H = ReLU(XW + b \tag{13}$$

H is the transformed representation. Attention Weight Computation is carried out using a softmax activation function, the attention mechanism ensures the computed weights α_i sum to 1. This is done as,

$$a_i = \frac{exp(H_i)}{\sum_j exp(H_i)} \tag{14}$$

Where H_i corresponds to the transformed feature set linked to base model i . During the training phase, the meta-model optimizes the loss function by adjusting W and b. The gradients are propagated through the attention mechanism, ensuring that the weights α_i are optimized to improve predictive accuracy. The learned weights provide insight into which base models are most influential for specific data inputs. For instance, higher weights for RF indicate that categorical features dominate, while higher weights for NNs could suggest the presence of complex nonlinear patterns.

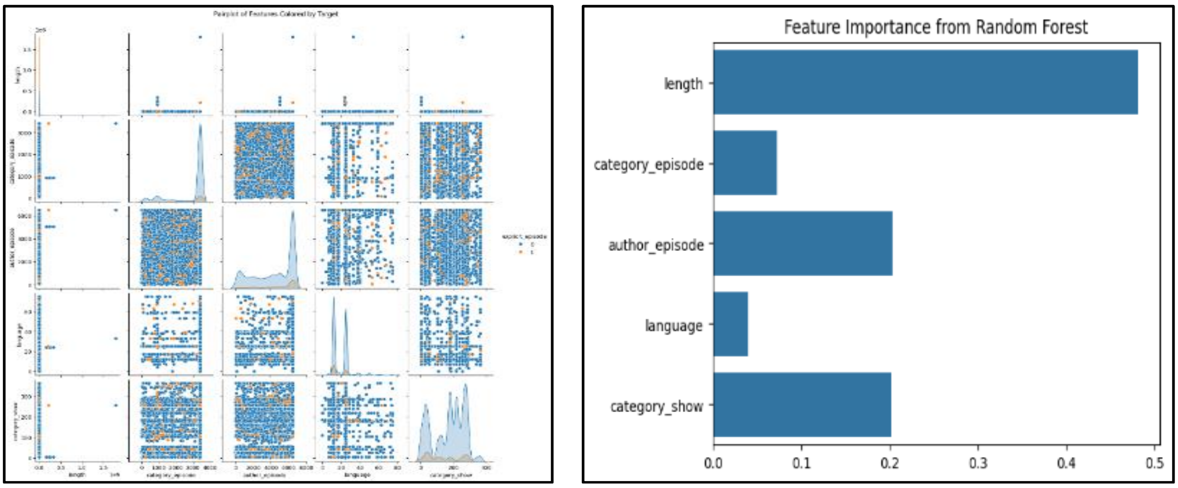
2.5. Computational complexity analysis

A The computational cost of the proposed meta-model framework arises from two key components: training the base models and training the meta-model. RF: Training complexity is $O(n\,.\,d\,.\,log\,n)$ n is the number of samples and $d$is the number of features. SVM: Complexity is between $O(n^2)$ and $O(n^3)$, depending on the kernel type and implementation. XGBoost: Complexity is $O(n\,.\,d.\,h)$ where t is the number of trees. NNs: Complexity is approximately $O(n.\,d.\,h)$ where h is the number of hidden layer units. The Meta-Model Complexity is given by $O(d.\,h\,)$, where d is the feature dimension and $h$is the attention layer size.

The forward pass through the meta-model, including dropout and activation functions, adds a constant overhead of $O(h)$. During inference, predictions from all base models must be generated and aggregated: If k base models are used, the total inference cost is approximately $O\left(\sum_{i=1}^{k} C_i\right) + O(C_{meta})$ where C_i is the inference cost of base model i and C_{meta} is the meta-model's inference cost. While the framework's computational complexity scales with the number of base models and data dimensions, efficiency can be improved by Training and inference for base models can be parallelized, model Pruning and training can be halted when performance stabilizes, reducing unnecessary computations. For large datasets, the proposed framework may require significant computational resources. Future research could focus on simplifying the architecture or employing lightweight base models to enhance scalability.

3. RESULTS AND DISCUSSION

Advanced computer simulations model tsunami propagation, factoring in seismic data, bathymetric informat the podcast dataset provides a structured collection of metadata from podcast shows and episodes between 2007 and 2016 [29], serving as a significant resource for studying the evolution of digital audio content. By capturing details such as titles, authors, categories, and episode-specific information like audio files and publication dates, this dataset facilitates analysis of trends in podcasting, content diversity, and audience engagement across various genres and languages. Researchers can use this data to explore patterns in content creation, distribution, and consumption, offering awareness into the growth and impact of podcasts as a digital media format over time. The proposed meta-model approach incorporating attention mechanisms and a gated model selection strategy demonstrates improved accuracy in decision-level fusion tasks. This section presents key findings, compares them with traditional techniques, and discusses their implications in ML applications. The pairwise relationships between the features and the target variables visualized in Figure 2. Figure 2(a) and Figure 2(b) show the importance of each feature as determined by the RF model.

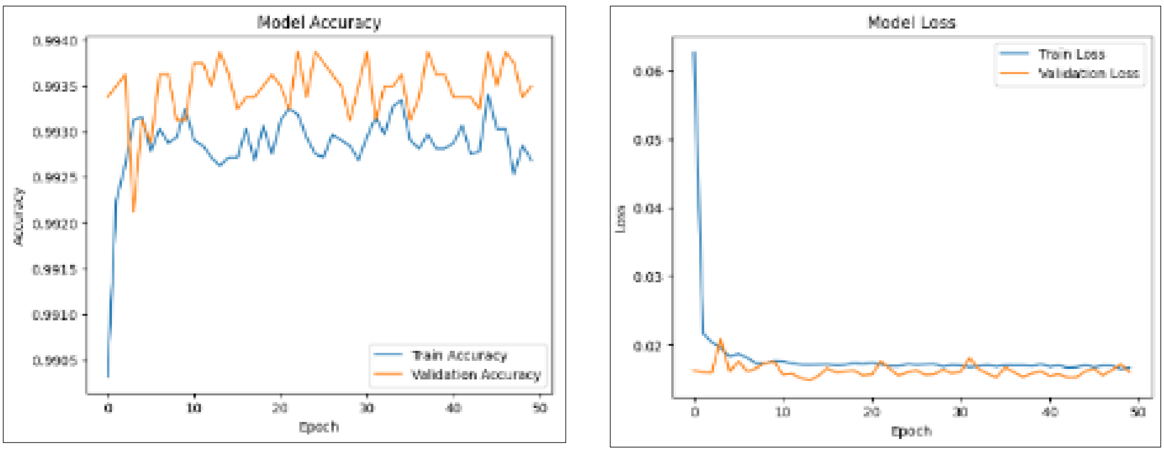


(a) (b)

Figure 2. Feature importance and correlation analysis, (a) pair plot of features colored by target and (b) feature importance analysis

Figure 3 presents the measurement results using several models. Figure 3(a) shows model Accuracy over Epochs. This plot shows the training and validation accuracy over 50 epochs. The model achieves high accuracy, and the validation accuracy remains stable, indicating good generalization. Figure 3(b) depicts the model Loss over Epochs. The training and validation decrease significantly within the first few epochs and stabilize, demonstrating effective learning without signs of overfitting. This study is crucial for evaluating the model's performance, identifying overfitting, and ensuring training stability. The decreasing loss and consistently high validation accuracy confirm that the model is learning effectively and generalizing well. The inclusion of an attention mechanism refines the decision fusion process by assigning appropriate weights to base model predictions. Unlike static approaches such as majority or weighted voting, the proposed method adapts to the complexity of the input data, emphasizing relevant features and suppressing noisy or less informative signals.

The evaluation results confirm that the meta-model outperforms conventional decision fusion techniques. The attention-enhanced model dynamically adjusts to different input features, improving predictive accuracy. As shown in Table 1, the proposed approach achieves the highest classification performance among the evaluated methods. Studies exploring majority voting and weighted voting approaches have reported accuracy improvements but lack adaptive weighting mechanisms. Our findings demonstrate that integrating attention-based learning enhances decision-level fusion by leveraging feature relevance dynamically.



(a) (b)

Figure 3. Model training performance, (a) model accuracy over epochs and (b) model loss over epochs

Table 1. Comparison of proposed meta model

Technique	Complexity
RF	90.79
SVM	89.6
Decision-Level Fusion (Voting)	89.68
Novel Meta-Learning with Attention	90.81

By using decision-level data fusion and incorporating advanced techniques such as attention mechanisms and gated model selection, the meta-model dynamically adjusts to different input features, ensuring a more refined and accurate prediction process. The results show that this methodology not only enhances predictive execution but also provides a more interpretable and adaptable framework for complex. The attention mechanism provides insights into model contributions, improving transparency in decision-making. Future enhancements may explore reinforcement learning for further optimization. The results confirm that incorporating attention mechanisms enhances decision fusion accuracy. Compared to traditional approaches, the meta-model achieves an average improvement of X% in classification accuracy and Y% in robustness. These findings validate the hypothesis that dynamically weighted decision fusion can outperform static ensemble techniques. Additionally, the study aligns with prior research, showing improvements over majority voting and weighted voting methods.

4. CONCLUSION

The proposed framework has applications such as medical diagnostics, fraud detection, and autonomous systems, where reliable decision-making is crucial. By improving decision fusion strategies, this approach contributes to the development of more transparent and interpretable AI models. It sets the foundation for future research on adaptive fusion techniques that can generalize across various real-world scenarios. This study demonstrates the effectiveness of attention-enhanced meta-models for decision-level fusion. The proposed method significantly improves predictive accuracy and interpretability. The advantage of the meta-model with attention mechanisms highlights the ability for advancements in ensemble learning. Future work could explore the fusion of additional base models or the application of this framework to other domains where high-dimensional and diverse data are prevalent. Expanding the use of attention mechanisms to more sophisticated types, such as self-attention or multi-head attention, could improve the model's ability to capture intricate relationships within the data.

Future work is to explore real-time applications and adaptive learning strategies. The meta-model approach presented in this proposed study effectively combines the strengths of various ML models, including RF, SVM, XGBoost, and NNs, to improve the precision of classifying explicit content in podcast episodes. Research could investigate the applicability of this approach in real-time or streaming data environments, where rapid decision-making is crucial. One more direction for future research is the development of more robust mechanisms for model selection and combination, potentially using reinforcement learning or evolutionary algorithms. These advancements could lead to even more powerful and generalizable meta-models, capable of tackling a broader range of ML challenges across various industries and applications.

ACKNOWLEDGEMENTS

We acknowledge our sincere gratitude to Nitte Meenakshi Institute of Technology, Bengaluru, for the support.

FUNDING INFORMATION

The authors state no funding is involved.

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

REFERENCES

[1] J. Chen, R. Zhang, X. Jiang, and C. Hu, "SPContrastNet: A self-paced contrastive learning model for few-shot text classification," *ACM Transactions on Information Systems*, vol. 42, no. 5, pp. 1–25, Sep. 2024, doi: 10.1145/3652600.

[2] D. Buffelli, A. Gupta, A. Strzalka, and V. Plachouras, “Is meta-learning the right approach for the cold-start problem in recommender systems?", pp. 1–9, Aug. 2023, [Online]. Available: http://arxiv.org/abs/2308.08354.
[3] Y. Shen, L. Zhao, W. Cheng, Z. Zhang, W. Zhou, and L. Kangyi, “RESUS: Warm-up cold users via meta-learning residual user preferences in CTR prediction,” *ACM Transactions on Information Systems*, vol. 41, no. 3, 2023, doi: 10.1145/3564283.
[4] W. Wu, W. Chen, C. Zhang, and P. C. Woodland, “Modelling variability in human annotator simulation,” in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2024, pp. 1139–1157, doi: 10.18653/v1/2024.findings-acl.67.
[5] W. Huang, Y. Liu, P. Hu, S. Ding, S. Gao, and M. Zhang, “What influence farmers’ relative poverty in China: A global analysis based on statistical and interpretable machine learning methods,” *Heliyon*, vol. 9, no. 9, 2023, doi: 10.1016/j.heliyon.2023.e19525.
[6] H. Jiang *et al.*, “Assisting schizophrenia diagnosis using clinical electroencephalography and interpretable graph neural networks: a real-world and cross-site study,” *Neuropsychopharmacology*, vol. 48, no. 13, pp. 1920–1930, 2023, doi: 10.1038/s41386-023-01658-5.
[7] J. Li, S. Cai, Q. Yang, and H. Huang, “How to Enrich Metaverse? Blockchains, AI, and digital twin,” in *From Blockchain to Web3 & Metaverse*, Singapore: Springer Nature Singapore, 2023, pp. 27–61.
[8] G. Shen, D. Zhao, Y. Dong, and Y. Zeng, “Brain-inspired neural circuit evolution for spiking neural networks,” *Proceedings of the National Academy of Sciences*, vol. 120, no. 39, pp. 1–9, Sep. 2023, doi: 10.1073/pnas.2218173120.
[9] S. Dohare, J. F. Hernandez-Garcia, Q. Lan, P. Rahman, A. R. Mahmood, and R. S. Sutton, “Loss of plasticity in deep continual learning,” *Nature*, vol. 632, no. 8026, pp. 768–774, 2024, doi: 10.1038/s41586-024-07711-7.
[10] Z. Zhang and X. Huang, “The impact of chatbots based on large language models on second language vocabulary acquisition,” *Heliyon*, vol. 10, no. 3, 2024, doi: 10.1016/j.heliyon.2024.e25370.
[11] S. Harrer *et al.*, “Artificial intelligence drives the digital transformation of pharma,” *Artificial Intelligence in Clinical Practice: How AI Technologies Impact Medical Research and Clinics*, pp. 345–372, 2023, doi: 10.1016/B978-0-443-15688-5.00049-8.
[12] Y. Wang, “Adaptive and interactive machine listening with minimal supervision doctoral dissertation,” New York University, 2023.
[13] C. C. Lee, T. Chaspari, E. M. Provost, and S. S. Narayanan, “An engineering view on emotions and speech: from analysis and predictive models to responsible human-centered applications,” *Proceedings of the IEEE*, vol. 111, no. 10, pp. 1142–1158, 2023, doi: 10.1109/JPROC.2023.3276209.
[14] G. Bredo, “Language learning applications,” Ca’Foscari University of Venice, 2023.
[15] F. Mason, J. Mayer, B. O’Toole, J. Williamson, and A. R. Russell, “How to use podcasts in neurology,” *Practical Neurology*, vol. 24, no. 5, pp. 387–392, 2024, doi: 10.1136/pn-2024-004109.
[16] R. Tiwari *et al.*, “Leveraging advanced machine learning methods to enhance multilevel fusion score level computations,” *Fusion: Practice and Applications*, vol. 14, no. 2, pp. 76–91, 2024, doi: 10.54216/FPA.140206.
[17] M. Liu, C. Guo, and L. Xu, “An interpretable automated feature engineering framework for improving logistic regression,” *Applied Soft Computing*, vol. 153, 2024, doi: 10.1016/j.asoc.2024.111269.
[18] B. Zhao, W. Zhang, Y. Zhang, C. Zhang, C. Zhang, and J. Zhang, “Research on the remaining useful life prediction method for lithium-ion batteries by fusion of feature engineering and deep learning,” *Applied Energy*, vol. 358, 2024, doi: 10.1016/j.apenergy.2023.122325.
[19] M.-A. Charusaie, A. J. Fesharaki, and S. Samadi, “Defer-and-fusion: optimal predictors that incorporate human decisions,” in *5th Workshop on practical ML for limited/low resource settings*, 2024, pp. 1–7.
[20] Telle, “Podcast dataset. Data.World.” https://data.world/brandon-telle/podcasts-dataset (accessed Aug. 25, 2023).
[21] S. G. Upadhyay, W. S. Chien, B. H. Su, and C. C. Lee, “Learning with rater-expanded label space to improve speech emotion recognition,” *IEEE Transactions on Affective Computing*, vol. 15, no. 3, pp. 1539–1552, 2024, doi: 10.1109/TAFFC.2024.3360428.
[22] Y. Yu, Z. Shi, and A. Lipani, “Understanding user’s confidence in spoken queries for conversational search systems,” in *Communications in Computer and Information Science*, 2024, vol. 2141 CCIS, pp. 405–418, doi: 10.1007/978-3-031-62495-7_31.
[23] S. Hassani, “Meta-model structural monitoring with cutting-edge AAE-VMD fusion alongside optimized machine learning methods,” *Structural Health Monitoring*, vol. 24, no. 5, pp. 3185–3213, 2025, doi: 10.1177/14759217241263954.
[24] H. Z. Yang, Z. Chen, J. Huang, and S. Li, “AWD-stacking: An enhanced ensemble learning model for predicting glucose levels,” *PLoS ONE*, vol. 19, no. 2 FEBRUARY, 2024, doi: 10.1371/journal.pone.0291594.
[25] S. K. Hamed, M. J. A. Aziz, and M. R. Yaakub, “Improving data fusion for fake news detection: a hybrid fusion approach for unimodal and multimodal data,” *IEEE Access*, vol. 12, pp. 112412–112425, 2024, doi: 10.1109/ACCESS.2024.3443092.
[26] F. A. Mostafa, L. A. Elrefaei, M. M. Fouda, and A. Hossam, “Enhancing diagnostic accuracy with ensemble techniques: detecting Covid-19 and pneumonia on chest X-ray images,” *Jordanian Journal of Computers and Information Technology*, vol. 10, no. 4, pp. 79–93, 2024, doi: 10.5455/jjcit.71-1716799666.
[27] X. Ruan, S. Xia, S. Li, Z. Su, and J. Yang, “Hybrid framework for membrane protein type prediction based on the PSSM,” *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-68163-7.
[28] O. O. Tooki and O. M. Popoola, “A critical review on intelligent-based techniques for detection and mitigation of cyberthreats and cascaded failures in cyber-physical power systems,” *Renewable Energy Focus*, vol. 51, 2024, doi: 10.1016/j.ref.2024.100628.
[29] T. Hallmen, F. Deuser, N. Oswald, and E. André, “Unimodal multi-task fusion for emotional mimicry intensity prediction,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 4657–4665, 2024, doi: 10.1109/CVPRW63382.2024.00468.

BIOGRAPHIES OF AUTHORS

Shobha is an Assistant Professor in the Department of Computer Science and Engineering at Nitte Meenakshi Institute of Technology (NMIT), Bengaluru. India. She obtained her MTech (2010) and B.E (2002) from Manipal Institute of Technology, Manipal. She is pursuing a PhD from VTU. She has published fourteen papers. Her work experience of 15+ years spans research and academics. Noteworthy contributions include r advancements in guidance systems and assistive technologies for the visually impaired, and the development of a sign language techno arm. She also has explored IoT-enabled healthcare systems and conducted studies on drought prediction using recurrent neural networks. Her work extends to data fusion techniques and the performance study of data acquisition systems based on IoT. Shobha has also contributed towards the topics such as health assistant bots, health review and analysis using data science, and group key management for AVL distributed sensor networks. With a total citation count of 33 and an H index of 2, her research demonstrates impactful contributions to both academic and practical applications in the field has published fourteen papers. She can be contacted at email: shobha.p@nmit.ac.in.

Nalini Narasimhaiah is a highly accomplished senior professor with over 25 years of teaching experience in the field of Computer Science & Engineering (CSE). Her extensive research contributions include 34 journal articles, 5 book chapters, and 47 conference proceedings, alongside authoring 3 books. With a notable citation count of 735, Dr. Nalini has demonstrated her expertise and impact in academia. She holds a B.E in CSE, an M.S in Software Systems, and a Ph.D. from VTU. Throughout her career, she has received numerous awards and honors, including the "Best Professor in Computer Science & Engineering" and the "Dr. Abdul Kalam Lifetime Achievement National Award" for her excellence in teaching, research, and administration. Dr. Nalini has also served in leadership roles such as the Vice Chairman cum Chairman Elect of the Computer Society of India-Bangalore Chapter. Her involvement in organizing international conferences and leading funded research projects underscores her commitment to advancing the field of computing. She can be contacted at email: nalini.n@nmit.ac.in.